157
12-8-7

VOLUME 66
NUMBER 1 to 17

WHOLE NO. 333 to 349
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

47842/7

Studies in Reputation:

I. Sex and Grade Differences in School Children's Evaluations of Their Peers

II. The Diagnosis of Social Adjustment

By

READ D. TUDDENHAM
University of California

LIBRARY

Accepted for publication, January 25, 1951

Price $1.50

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

[illegible] Research
[illegible]
12·8·70
J 157

ACKNOWLEDGMENTS

THE data of this investigation are from the Guidance Research Study, which is being carried on at the University of California Institute of Child Welfare under the direction of Professor Jean W. Macfarlane. Many members of the Guidance Research Study staff have contributed to this research, but special acknowledgments are due to Mrs. M. H. Davis and Mrs. Loretta Maxfield who collected the data, to Dr. Marjorie Honzik who did a preliminary study, and to Mr. Jack Delaney who rendered invaluable assistance in the statistical analysis of the data. The writer takes this opportunity to thank Dr. Macfarlane, Dr. Harold E. Jones, and Dr. Herbert S. Conrad for encouragement and criticism in the preparation of this manuscript.

Clerical and statistical support has been supplied by the WPA and by the U. S. Public Health Service. The Guidance Research Study proper was made possible by funds from the Laura Spelman Fund, General Education Board, Rockefeller Foundation, and the University of California.

READ D. TUDDENHAM

47821

TABLE OF CONTENTS

I. SEX AND GRADE DIFFERENCES IN SCHOOL CHILDREN'S EVALUATION OF THEIR PEERS

I. *Introduction: The Reputation Test* 1

II. *Population and Methods of Administration* 5

III. *Patterns in Reputation* 9

IV. *Patterns in Judgment* 20

V. *Agreement among Judges on the Reputation Test* 30

VI. *Summary and Conclusions* 37
References 58

II. THE DIAGNOSIS OF SOCIAL ADJUSTMENT

I. *The Diagnostic Use of the Reputation Test* 40

II. *Illustrative Case Records* 45

I. SEX AND GRADE DIFFERENCES IN SCHOOL CHILDREN'S EVALUATION OF THEIR PEERS

CHAPTER I

INTRODUCTION: THE REPUTATION TEST

A. PREVIOUS INVESTIGATIONS

IN STUDYING child personality, most investigators utilizing rating procedures have depended upon the impressions of adults—parents, teachers, or psychologists. However, the influence of other children constitutes a major component of the social milieu to which the child must adapt, and his behavior with them may differ in many ways from that which he exhibits in other groups. This important domain is peculiarly inaccessible to the adult observer, whose very presence alters the situation under study. In order to investigate those aspects of child personality which are revealed in the social relationships obtaining among children, a method is needed which permits the children themselves to express their attitudes toward one another in a manner which is adapted to their capacities, which avoids disturbing them or provoking undesirable attitudes of "tattling" on their associates, and which meets the practical criteria of ease of administration, quantifiability, and reliability of measurement.

To meet this need, Hartshorne, May, and Maller in 1929 devised the "Guess Who" test in connection with their investigations of service and altruism among school children (3). This test consisted of a series of short word pictures presented with the instructions that the children "guess" which of their classmates were most nearly like the persons described. This procedure successfully circumvented the practical difficulties of securing a measure of reputation, and achieved a high degree of reliability in the final ratings by pooling the extreme judgments of a large number of judges.

The method has subsequently been used by C. M. Tryon (9), by H. E. Jones (4), by Symonds and Jackson (8), and by Elliot (2) to study children's social adjustment; by Campbell (1) to study sex differences in judgment patterns of adolescents; and by Tryon (10) to study adolescent personality values. Moreno (7) has adapted a similar procedure to the study of spontaneous social groupings. Results of these investigations have indicated that reputation testing offers a promising approach to a variety of problems in personality development, and further, that the method supplies a type of information which is not obtainable by means of the more common devices which secure ratings by adult observers.

For these reasons, reputation testing was added to the program of data collection of the University of California Guidance Study, and was administered beginning at the first-grade level (5). Since the subjects of this study were considerably younger at the beginning of the program than any who had been tested previously, modifications of item-selection, terminology, and mode of administration were required. Macfarlane, Honzik, and Davis (6) have published a preliminary report on the effectiveness of the method with these primary grade children. The present monograph is based on the complete data of this testing program at grade levels I, III, and V.

B. PURPOSE OF THE PRESENT INVESTIGATION

The purpose of this study is to explore the possibilities of the Reputation Test as a device for studying social aspects of personality development in children of elementary school age.

Analysis of the mass of data obtained from more than 2,700 individual test administrations has suggested many more avenues of research than could be dealt with in one monograph. The present reports have been restricted to two interrelated problems. On the normative level, we have studied group differences in reputation as related to age and sex for the light they throw on the general process of social development during this period of childhood. On the level of qualitative diagnosis, we have investigated the usefulness of the Reputation Test in locating the socially atypical, disliked, or withdrawn children who need special help, by means of a comparison of "blind" interpretations of individual records with relevant data from other sources.

In a domain in which speculations are more numerous than facts, the supplying of normative data and the refinement of method are necessary preliminaries to the more crucial tests of theory. At the risk of violating a major criterion of scientific respectability, the present investigation has been formulated not as a statistical proof or disproof of any explicit hypothesis, but rather as an exploration of methods for studying this area of personality and its developmental sequences.

C. DESCRIPTION OF THE REPUTATION TEST

The Reputation Test, as used in this study, consisted of a series of brief word pictures to each of which the child was asked to respond with the name of the persons fitting the description.

Following the usage of Tryon (10) and Symonds and Jackson (8), the items of the test were arranged in pairs of opposites, one designating the positive or desirable pole of behavior, the other the undesirable pole along the same dimension. For example, items 13 and 14 read: (—) "Which children are poor sports? (are poor losers?)" and (+) "Which children are good sports? (good losers?) (play fair?)." The items were originally categorized as favorable or unfavorable for adjustment on the basis of clinicians' judgments. It was later discovered that all the "favorable" items were positively correlated with the item hereafter referred to as POPULAR. Throughout the discussion the individual questions will be referred to interchangeably as "items" or "traits"; paired questions will be referred to interchangeably as "item-pairs," "dimensions," and "variables."

The test as used in the first, second, and certain third grades consisted of ten pairs of items referring to either sex, a pair of items referring exclusively to boys and a pair to girls, and one unpaired item, "Who is your best friend?" Three additional item-pairs were added in the testing of certain third- and all fourth-grade classes, since the older children were able to handle a longer test without fatigue. Four more pairs were added when the test was given to fifth- and sixth-grade classes.

For children in the primary grades it was considered essential that the wording of items be as brief and simple as possible. The expression, "Here is someone who is—" used by Tryon, was abandoned in favor of a direct question, and the vocabulary chosen from words actually used by the children in interviews. At grade levels V and VI, some items were lengthened where experience had shown that additional descriptions were advisable, and the vocabulary was adapted to the increased age of the subjects, but the original wording of the items was not materially altered in most instances. The items TAKES CHANCES, REAL BOY, and ACTS LIKE LITTLE LADY and their antonyms were rephrased because experience suggested the possibility that the original formulas might no longer be conveying the intended meaning.[1] The original and revised forms of the items are presented in Tables 1 and 2, respectively, together with the titles and symbols which will be used for the sake of brevity in referring to them in subsequent pages. The titles appended in the course of data analysis are not in all cases synonymous with the actual items to which the children responded and upon which all findings are based. This fact should be kept in mind in interpreting the contents of following chapters.

---

[1] Despite the changes in phrasing, these items showed the same regular trends with age as did the unmodified items. See Chapter III.

TABLE 1

THE ORIGINAL FORM OF THE REPUTATION TEST (INDIVIDUAL ADMINISTRATION)

Item	Title	Symbol	Item
− 1.	WIGGLY		Which children wiggle a lot and can't sit still?
+ 2.	QUIET	Q	Which children sit very still and quiet?
+ 3.	POPULAR	P	Who are the ones everyone likes?
− 4.	NOT MANY FRIENDS		Who are the ones nobody likes very much?
+ 5.	FULL OF FUN	FF	Which children are always smiling and laughing? (are full of fun?)
− 6.	SERIOUS		Which children don't smile very much and seem sort of sad?
− 7.	QUARRELSOME		What children quarrel a lot? (get into arguments?)
+ 8.	NOT QUARRELSOME	Ql−	What children hate to quarrel? (don't quarrel much and keep out of arguments?)
− 9.	DOESN'T TAKE CHANCES		Which children get scared of everything? (are 'fraidy cats?) (are too scared to take chances?)
+ 10.	TAKES CHANCES	C	Who are the bravest and almost never get scared? (take big chances and are *not* easily scared?)
− 11.	BOSSY		What children are bossy? (try to run things?)
+ 12.	NOT BOSSY	B−	Which children let other children boss them? (usually give in?)
− 13.	POOR SPORT		Which children are poor sports? (are poor losers?)
+ 14.	GOOD SPORT	S	Who are good sports? (good losers?) (play fair?)
− 15.	BASHFUL		Which children are the most bashful? (hate reciting?)
+ 16.	NOT BASHFUL	Ba−	Which children aren't the least bit bashful? (don't mind reciting?)
+ 17.	GOOD AT GAMES	G	Which children are the best at outdoor games?
− 18.	NOT GOOD AT GAMES		Which children aren't very good at games?
− 19.	GETS MAD EASILY		Which children get mad the easiest?
+ 20.	DOESN'T GET MAD	M−	Which ones don't get mad much?
− 21.	SISSY		What boys are the worst sissies?
+ 22.	REAL BOY	RB	Who are the real boys—the regular fellows? (are never sissies—they just act like boys?)
− 23.	TOMBOY		What girls are tomboys? (act like boys and play boys' games?)
+ 24.	ACTS LIKE LITTLE LADY	LL	What girls act like little ladies?
+ 25.	BEST FRIEND	BF	Who's your best friend?

Note: See Table 2 for the additional item-pairs GOOD-LOOKING—NOT GOOD-LOOKING, SHOW-OFF—NOT A SHOW-OFF, and LEADER—NOT A LEADER, which were added to the test administered by interview to 11 of the 20 third-grade classes and to fourth-grade classes.

TABLE 2

THE REVISED FORM OF THE REPUTATION TEST (GROUP ADMINISTRATION)

Item	Title	Symbol	Item
− 1.	Wiggly		Which children wiggle a lot and can't sit still, or get up and walk around a lot?
+ 2.	Quiet	Q	Which children sit very still and quiet?
+ 3.	Popular	P	Who are the ones everybody likes?
− 4.	Not Many Friends		Who are the ones nobody seems to care much about?
+ 5.	Full of Fun	FF	Which children are always smiling and laughing and are full of fun?
− 6.	Serious		Which children don't smile much and seem sort of sad or serious?
− 7.	Quarrelsome		Which children quarrel and argue a lot?
+ 8.	Not Quarrelsome	Ql −	Which children don't quarrel much and usually keep out of arguments?
+ 10.	Takes Chances	C	Who are the ones that like to take big chances when they are playing games?
− 9.	Doesn't Take Chances		Which ones hardly ever take chances and don't like to do it?
− 11.	Bossy		Which children are bossy and try to run things?
+ 12.	Not Bossy	B −	Which children give in to others and hardly ever boss?
+ 14.	Good Sport	S	Who are the good sports, the ones that always play fair?
− 13.	Poor Sport		Which ones aren't very good sports, the poor losers?
− 15.	Bashful		Which children are the most bashful and don't like to recite in class before visitors?
+ 16.	Not Bashful	Ba −	Which children aren't the least bit bashful and don't mind reciting in class before visitors?
+ 17.	Good at Games	G	Which children are good at games and play them well?
− 18.	Not Good at Games		Which children aren't very good at games?
− 19.	Gets Mad Easily		Which children get mad the easiest, lose their temper?
+ 20.	Doesn't Get Mad	M −	Which children hardly ever get mad or lose their temper?
+ 22.	Real Boy	RB	Who are the real boys that enjoy doing all the things boys do?
− 21.	Sissy		What boys *don't* enjoy doing all the things regular boys do?
− 23.	Tomboy		What girls like to play boys' games and do the things boys do?
+ 24.	Acts Like Little Lady	LL	What girls *don't* like boys' games at all and just do the things girls do?
+ 26.	Good-Looking*	GL	Who are the ones everybody thinks are good looking?
− 27.	Not Good-Looking		Who are the ones very few people think are good looking?
− 28.	Show-Off		Which ones show off a lot and act silly?
+ 29.	Not a Show-Off	Sh −	Which ones hardly ever show off or act silly?
+ 30.	Leader	L	Which children are good at starting games and getting things going—the ones that suggest things to do?
− 31.	Not a Leader		Who are the ones that just sit around and never start anything?
+ 32.	Tidy**	T	Which children look most neat and tidy?
− 33.	Untidy		Which children don't look very neat and tidy?
34.	Fights		Which children get into lots of fights and like to fight?
+ 35.	Doesn't Fight	F −	Which children never get into fights and hate to fight?
+ 36.	Friendly	Fr	Which children are very friendly and nice to everybody?
− 37.	Not Friendly		Which children don't care much about making friends or being friendly?
+ 38.	Takes Joke on Self	J	Which children can laugh at a joke on themselves?
− 39.	Can't Take a Joke		Which children can never laugh at a joke on themselves—they just get mad or cry about it?
+ 25.	Best Friend	BF	Who is your best friend?

* Items 26–31 were added for 11 of the 20 third-grade classrooms and for the fourth grade.

** Items 32–39 were added for fifth- and sixth-grade classrooms.

Chapter II

POPULATION AND METHODS OF ADMINISTRATION

A. Population Studied

The present investigation is based upon Reputation Test data collected as a part of the University of California Guidance Study, a longitudinal investigation of the physical, psychological, and social development from birth to maturity of 250 Berkeley children. The method of selection and demographic characteristics of the Guidance Study sample have been described in detail by Macfarlane (5). Suffice it to note here that they constitute a subsample from a group consisting of every third child born in Berkeley, California, between January 1928, and July 1929, and that their families resembled the general urban population of the United States (1930 census), though differing in some respects. Since the Guidance Study files afford rich and varied psychological material on the subjects from sources other than the Reputation Test, all individual cases used to illustrate the diagnostic value of the Reputation Test were drawn from the Guidance Study sample. (See "Studies in Reputation. II. The Diagnosis of Social Adjustment.")

The normative studies of reputation reported in subsequent chapters were based on a much larger group than that described above. The Reputation Test can be given only to a whole classroom. The need to secure data periodically on as many Guidance Study cases as possible within existing time limitations determined the choice of schoolrooms to be tested. However, the classmates of Guidance Study subjects constituted a very much larger proportion of the final population than did the Study cases themselves. The Guidance Study children were by age six scattered widely throughout all the school districts of Berkeley, and owing to special promotions and failures, soon came to represent a grade range of three or more years. Since only a few of the classrooms in which Guidance Study cases were located could be tested in any one year, the program of data collection was of necessity irregular. However, the sampling of the Berkeley school population was a broad one, and all elementary schools in Berkeley were represented at least once during the years the data collection was in progress.

The Reputation Test was administered in 85 classrooms including all six grade levels of the elementary school system. In order to simplify statistical treatment, results from classrooms of the same grade level were pooled and data analyzed only for the alternate grade levels I, III, and V, a procedure which was adequate to reveal the trends with age which constituted on focus of our interest. Shifts in reputation within the same classroom over a long temporal interval were not studied because retesting had not been sufficiently extensive or regular. Socioeconomic differences among classrooms pooled were disregarded, but the size of the total sample and the nonsystematic selection of classrooms for inclusion at any level served to reduce sampling errors arising from this treatment. Future research might profitably be directed to these problems of temporal change in individual reputation and of differences in judgment patterns at various socioeconomic levels.

The number of children on whom the present report is based is given in Table 3. It should be understood that

TABLE 3

NUMBER OF CHILDREN FOR WHOM REPUTATION TEST DATA WERE ANALYZED

Grade	No. of Classes	Number in Class			Number Present		
		Girls	Boys	Total	Girls	Boys	Total
First Grade Items 1 to 25	4	49	70	119	41	65	106
Third Grade Items 1 to 25	20	340	352	692	318	336	654
Items 26 to 31	11	178	194	372	169	184	353
Fifth Grade Items 1 to 39	18	306	322	628	296	312	608

all analyses referring to reputation scores are based on the "number in class" (including absentees), while all analyses referring to votes given or to self-mentions are based on the "number present."

B. ADMINISTRATION

In Grades I, II, III, and IV, the test was administered to each child in an individual interview. After the purposes of the test had been explained to the teacher and the principal, the teacher was asked to send to the interview room one child at a time, starting with the one or two socially most secure and popular, then the others in any order.

To the child, instructions were as follows:

> "I want to see how good you are at guessing the names of children in your room. I'll tell you just what these children do, and you tell me who I'm talking about. It may be one person or two or three. Sometimes it might even be you."

The questions were then asked of the child one by one. Whenever a child gave evidence of failing to understand the original phrasing of a question, the indicated alternate formulas were substituted. The interviewer recorded each child's nominations on a sheet prepared for each item-pair with the pupils' names listed down one side and repeated across the top. Individual mentions on the favorable or unfavorable item of a dimension were indicated respectively by a plus or minus sign placed in the column of the child mentioning, and the row of the child mentioned, a method which facilitated analysis both of votes received and of votes given. The teacher, also, was asked to supply information in the same form about the children in her class. By this procedure, data of four sorts—mentions of others, self-mentions, mentions by classmates, and mentions by the teacher—were obtained concurrently for each child.

The test took as a rule about eight or ten minutes for each child, the increasing length of the test from grade to grade being compensated for by the children's greater maturity and verbal facility.

For Grades V and VI, the extremely time-consuming method of individual administration was abandoned in favor of group administration to each class, after the feasibility of the change had been established by a comparison of individual vs. group test results from a sample room.[1]

---

[1] In one fifth-grade classroom, individual interviewing was followed after a two-month interval, by a written group testing. Although discrepancies might have arisen from changes in actual reputation between testings as well as from the effects of the change in administrative procedure, the two sets of data were in close agreement, a

The verbal instructions which preceded distribution of the test booklets were:

> "Here are some questions about things children do or the way they feel. Write after each question the names of any boys or girls in your class who may fit it. You may name yourself. If the question does not fit anybody in the class, skip it. Before you begin, be sure to put your name on the front cover of the booklet."

In the case of nonreaders, the examiners made a routine practice of reading the items to the children and recording their nominations.

Group testing reduced the time required per class of 30 from between four and five hours to approximately 40 minutes, and clearly justified itself by increasing correspondingly the number of classrooms which could be covered and by making possible the inclusion of additional items in the test. However, the change from individual to group administration rendered the results for upper grades not strictly comparable to those for lower grades. A more serious disadvantage was the loss of the spontaneous comments by the children, which had supplied a constant check on how the items were being interpreted. Among the many such remarks collected, the following illustrate this point:

> Acts Like Little Lady (from a girl): "Lucy acts like a little lady. She always carries a handkerchief and wears a coat."
>
> Popular (from a boy): "Oh, everybody likes Louise because she's so nice and rough. She's manager of our gang and I'm vice-manager."
>
> Poor Sport (from a girl): "I heard Bobby say, 'Oh, damn those old girls!' So he's a poor sport."

C. Attitudes Toward the Reputation Test

Children in the primary grades usually made a favorable response to the test as a whole. "I wish I could have two turns." "Your game was an awful lot of fun. Thanks a lot." "Is there a prize for the one that guesses the most right?" Occasionally in the fourth grade, and more frequently in the fifth and sixth grades of schools drawing from the middle and upper social strata, a child would reject an item, as for example: "I don't like to recommend anybody as a poor sport." Rejection, sometimes extending to the whole test, seems to be less infrequent at the junior high-school level where social acceptance becomes a matter of uneasy concern for many adolescents. It was not, in our experience, an obstacle in the elementary school.

A more serious problem was the attitude of some school administrators that the Reputation Test might make the children critical of each other, and undo school attempts to provide an acceptant and tolerant social atmosphere. At the beginning of the program of data collection, the Guidance Study staff shared this concern. Interviewers were told to watch for signs of possible discomfort in their informants, and to discontinue if they occurred. However, the alacrity and matter-of-factness with which nominations were volunteered soon made it clear that the Reputation Test was merely tapping attitudes already well crystallized in each classroom group. In

---

fact which suggested that reputational status has considerable constancy over this interval. The mean correlation between individual and group test scores on a single item was .77. The range was from .48 to .95, but the lower correlations tended to be more the consequence of few mentions and small SD's than of actual disagreement. On only one item was there a significant difference between mean scores obtained by the two testings. One trend did appear, however. Including nonsignificant differences, 17 of the 19 favorable items yielded more mentions in the individual interview, while 13 of the 18 unfavorable items yielded more mentions on the group test—an amusing reversal of the pattern of adults who are usually more willing to express critical judgments in speech than in writing.

requesting cooperation from school principals, we therefore attempted to anticipate and answer this objection, and to emphasize the very real value of the test to the school, in locating the children most in need of social support. The formulation used is quoted in full in Macfarlane's monograph (5, p. 53).

Parents were naturally interested in the testing program, but only a handful made complaints to the schools. Usually they were parents of insecure or unpopular children, and in nearly all instances were reassured when they learned of the schools' efforts to utilize the test results constructively to improve the social adjustment of unaccepted children.[2]

[2] See Elliott (2) and Tuddenham (11) for examples of practical applications of the results in this way.

CHAPTER III

PATTERNS IN REPUTATION

THE CAPACITY of the Reputation Test to reveal the opinions which a classroom group has formed concerning a given class member is one of its principal values in individual diagnosis, but the significance of a child's reputation can be assessed only in relation to appropriate norms. In the present chapter such norms are presented by sex and grade level for the Berkeley population tested. The results are of considerable intrinsic interest for the light they throw upon grade-school children's conceptions of masculine and feminine personality attributes, and upon changes in these conceptions with advancing grade level. The correspondence between the children's attitudes and the stereotypes of adults in American society at large suggests that the trends evidenced in the data may apply much more widely than just to the California group investigated.

A. STATISTICAL TREATMENT

In dealing with children's reputations with classmates, each pair of items was treated as a single dimension, a linear continuum whose extremes were defined by the opposed items of the pair concerned. On each such dimension, the number of mentions *received* by a given child on the favorable and unfavorable items of the pair were summed algebraically to yield a measure of his reputation on the dimension as a whole. Since it was desired to pool data from all the classrooms of the same grade, it was further necessary to correct the algebraic sums of votes received for variations in the size of the different classes. The corrected score for each child on each item-pair dimension was obtained by dividing the algebraic sum of plus and minus votes received by the number of children present in class for the test, exclusive of the child himself, and multiplying by 100. (Self-mentions were treated separately.) This percentage is referred to hereafter as the *item-pair score* or *dimension score*. A child mentioned by everyone else in his class as possessing the favorable attribute had a score of +100, while one mentioned by all as possessing the undesirable quality had a score of −100. The neutral point of nonmention (or of an equal number of plus and minus mentions) was zero, a score which proved to be modal on all dimensions, regardless of sex or grade level. This last finding is a consequence not only of the cancellation procedure used in scoring, but also of the nature of the test itself, which unlike most rating devices secures mentions only of children who are clearly extreme in the judgment of the rater. Anticipating the findings of Chapter V, roughly 20 per cent of the children in these age ranges received a zero score on the average item, and of these about half received a zero score in consequence of being named equally often for opposite traits. In most cases children in the latter group were mentioned not more than twice on the opposite poles of the dimension concerned. Thus, even though an algebraic summation procedure which tended to inflate the number of zero scores was used, the "indifference band" was quite narrow, and four out of five children received a definite characterization on the average item-pair.

A disadvantage of basing analysis upon item-pairs rather than items was the possibility that the items of a given pair might not be strict antonyms, and that cancellation of positive and negative votes on a given dimension might hence be inappropriate. While the items of a pair were phrased to be as "opposite" as the semantic acumen of the staff could make them, a further justification for this mode of treatment was the fact that in no instance did a child mention a given classmate on both items of a pair. A second possible disadvantage was the tendency of such algebraic summation to obscure small deviations from the indifference point. However, these small deviations are undoubtedly much less reliable than extreme scores and may not merit conservation. The statistical advantages of more nearly symmetrical score distributions and increased score variability were further reasons for treating the reputation data in terms of item-pairs rather than items.

B. Sex Differences in Scores

Means and standard deviations on the various dimensions of the Reputation Test are presented separately for boys and girls in Table 4. However, the essential findings are most easily grasped by inspection of Figure 1, which is drawn from the means presented in the table.

Perhaps the most striking fact to be noted is the marked difference in the reputation of the average girl and boy. Apparently, even first-grade children make rather clear-cut distinctions in the qualities they attribute to boys and to girls, associating quiet docility and friendliness with femininity—aggressiveness, boldness, and activity with masculinity.

Of equal importance is the fact that girls on the average enjoy *more favorable status* than boys on a majority of the item-pairs. In many instances this difference is highly significant (see Table 5). On certain item-pairs, the reputations of boys are not only less favorable than those of girls, but are indeed on the negative side of the indifference point, —the average boy being characterized as Quarrelsome, Bossy, Gets Mad Easily, Not Good-Looking, Wiggly, Untidy, and Fights. Evidence that the children themselves and not merely the test constructors regarded these items as unfavorable is supplied by their negative correlations with the dimension Popular–Not Many Friends in all sex-grade groups.

Few writers concerned with the sexual differentiation of adult personalities have emphasized the point, yet it seems clear that the marked difference in the climate of opinion faced by the average boy as compared with the average girl in relationships with other children could scarcely fail to produce a residual difference between men and women in social attitudes. We suspect that the atmosphere of disapproval surrounding the average small boy as contrasted with the small girl may well be a factor in the nonconformity and resistance to authority which are often attributed to male adolescents and adults in our society.

C. Grade Level Differences in Scores

The sex differences reported above suggest that stereotype thinking about "girl traits" and "boy traits" seriously influenced the judgments of our informants, even though these judgments were secured in the form of identifications of actual classmates on the various dimensions rather than descriptions of the "average" boy or girl. If, however, our findings reflect largely sex stereotypes of judges, we might expect these stereotypes to grow more differentiated with advancing grade level. In fact, exactly the converse was found, the mean reputation scores of the boys and of the girls tending to move closer together in successive grades. This trend seems to be primarily the consequence of a shift in the favorable direction for most of the

TABLE 4

MEANS AND STANDARD DEVIATIONS OF BOYS' AND GIRLS' SCORES ON THE VARIOUS ITEM-PAIRS OF THE REPUTATION TEST

Nos.	Titles*	Grade I				Grade III				Grade V			
		Girls (*N*=49)		Boys (*N*=70)		Girls (*N*=340, items 1–25) (*N*=178, items 26–31)		Boys (*N*=352, items 1–25) (*N*=194, items 26–31)		Girls (*N*=306)		Boys (*N*=322)	
		Mean**	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1– 2	QUIET	7.6	12.4	−3.9	14.5	5.2	13.9	−3.8	15.1	6.9	17.1	−4.3	16.6
4– 3	POPULAR	7.6	14.4	−1.1	15.4	4.7	14.3	− .8	14.9	4.8	20.7	3.0	17.3
6– 5	FULL OF FUN	5.9	11.2	1.3	10.3	2.4	10.3	1.3	9.0	3.2	12.1	3.2	11.4
7– 8	NOT QUARRELSOME	6.9	12.2	−5.7	10.9	3.9	9.6	−4.4	14.4	4.4	15.5	− .7	16.2
9–10	TAKES CHANCES	1.8	7.2	3.3	12.7	− .2	7.2	4.1	11.0	−1.9	10.4	3.9	15.2
11–12	NOT BOSSY	1.4	7.3	−2.9	9.6	1.3	10.0	−1.9	9.3	.8	13.1	− .9	14.5
13–14	GOOD SPORT	4.7	11.8	1.6	11.7	3.2	9.3	1.2	13.0	3.5	12.1	2.6	14.4
15–16	NOT BASHFUL	.6	7.1	2.6	8.7	1.3	9.0	2.1	7.5	.6	16.1	3.7	11.9
18–17	GOOD AT GAMES	4.9	10.9	.1	11.8	1.2	10.6	3.4	14.9	1.0	15.3	5.8	20.8
19–20	DOESN'T GET MAD	2.2	6.8	−1.4	9.9	2.4	10.5	−2.0	10.6	2.2	13.2	− .1	14.2
21–22	REAL BOY			2.6	16.1			4.6	16.7			8.6	23.2
23–24	LITTLE LADY	6.7	15.7			4.2	16.4			3.2	23.0		
25	BEST FRIEND	5.8	6.7	5.3	4.7	5.6	5.2	5.4	4.9	6.7	5.4	6.5	5.4
27–26	GOOD-LOOKING					3.6	13.0	− .5	6.9	3.0	19.7	− .5	12.7
28–29	NOT A SHOW-OFF					1.4	9.3	−1.9	9.9	2.8	12.9	−3.8	15.9
31–30	LEADER					1.0	5.0	1.9	9.7	.1	11.6	.9	13.7
33–32	TIDY									8.0	14.5	−1.2	14.4
34–35	DOESN'T FIGHT									2.8	8.6	−2.2	17.4
37–36	FRIENDLY									5.0	12.6	1.4	10.1
39–38	CAN TAKE A JOKE									1.3	8.7	2.3	8.9
	Mean (all dimensions)	4.7	10.3	.2	11.4	2.8	10.2	.6	11.2	3.1	13.8	1.5	14.4

---

* In this table, the titles refer to the favorable–unfavorable item-pairs of which the titles indicate the favorable poles.

** A Positive Mean indicates that the average score is on the favorable side of the indifference point.
A Negative Mean indicates the unfavorable side.

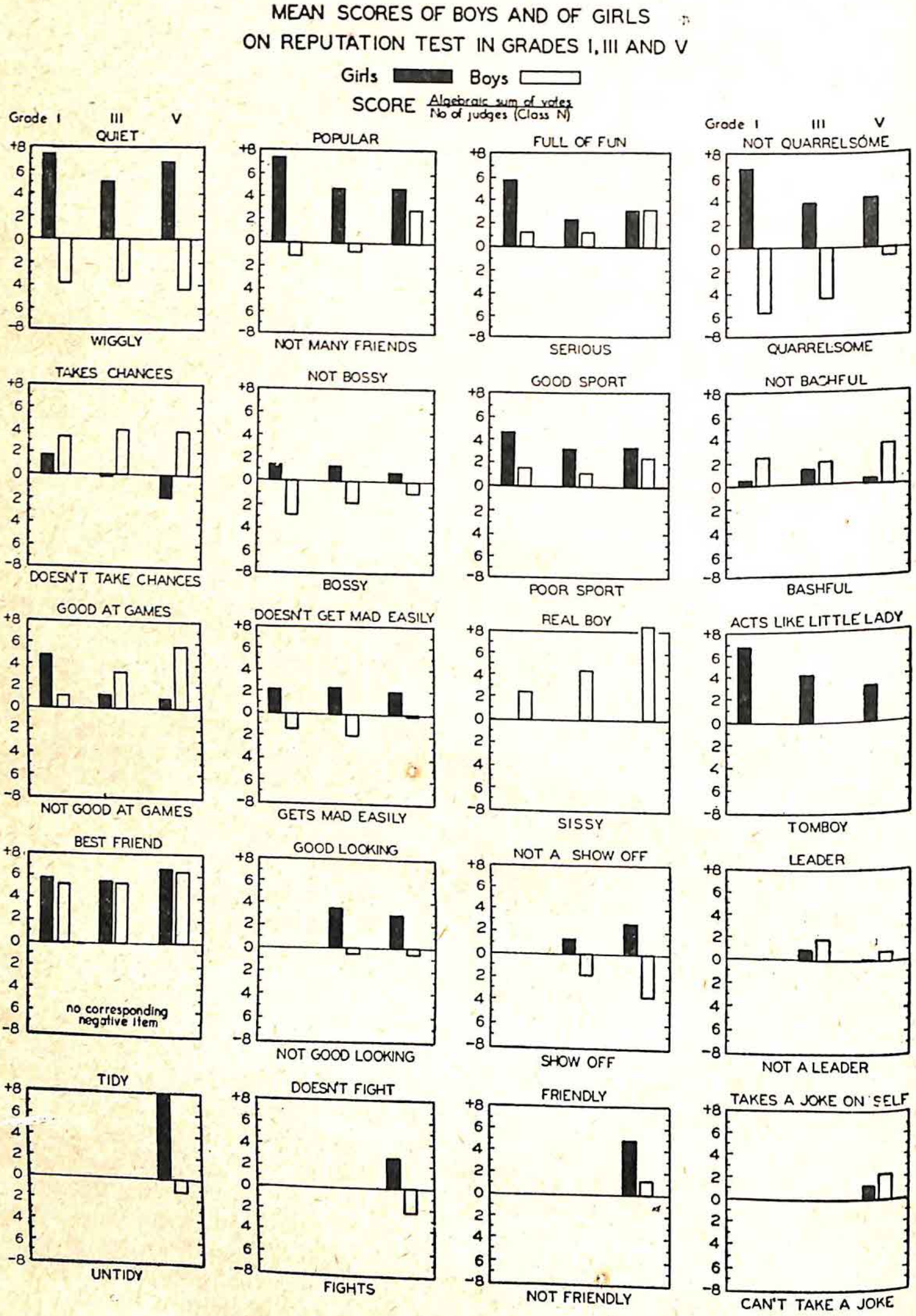

MEAN SCORES OF BOYS AND OF GIRLS
ON REPUTATION TEST IN GRADES I, III AND V
Girls
Boys
SCORE Algebraic sum of votes / No of judges (Class N)
Grade I III V
QUIET
WIGGLY
POPULAR
NOT MANY FRIENDS
FULL OF FUN
SERIOUS
NOT QUARRELSOME
QUARRELSOME
TAKES CHANCES
DOESN'T TAKE CHANCES
NOT BOSSY
BOSSY
GOOD SPORT
POOR SPORT
NOT BASHFUL
BASHFUL
GOOD AT GAMES
NOT GOOD AT GAMES
DOESN'T GET MAD EASILY
GETS MAD EASILY
REAL BOY
SISSY
ACTS LIKE LITTLE LADY
TOMBOY
BEST FRIEND
no corresponding negative item
GOOD LOOKING
NOT GOOD LOOKING
NOT A SHOW OFF
SHOW OFF
LEADER
NOT A LEADER
TIDY
UNTIDY
DOESN'T FIGHT
FIGHTS
FRIENDLY
NOT FRIENDLY
TAKES A JOKE ON SELF
CAN'T TAKE A JOKE
+8
6
4
2
0
2
4
6
-8

FIG. 1

TABLE 5

SIGNIFICANCE OF DIFFERENCES IN MEAN SCORES OF GIRLS VS. BOYS

Grade I		Grade III		Grade V	
Item-Pair	Critical Ratio	Item-Pair	Critical Ratio	Item-Pair	Critical Ratio
A. Item-Pairs On Which Girls Have More Favorable Reputation					
Not Quarrelsome	5.8	Not Quarrelsome	9.0	Quiet	8.3
Quiet	4.6	Quiet	8.1	Tidy	7.9
Popular	3.2	Doesn't Get Mad	5.5	Not A Show-Off	5.7
Not Bossy	2.8	Popular	4.9	Doesn't Fight	4.7
Doesn't Get Mad	2.4	Not Bossy	4.4	Not Quarrelsome	4.1
Full of Fun	2.3	Good-Looking	3.8	Friendly	4.0
Good at Games	1.8	Not a Show-Off	3.2	Good-Looking	2.6
Good Sport	1.4	Good Sport	2.4	Doesn't Get Mad	2.0
Best Friend	.4	Full of Fun	1.4	Not Bossy	1.6
		Best Friend	.6	Popular	1.2
				Good Sport	.8
				Best Friend	.6
B. Item-Pairs On Which Boys Have More Favorable Reputation					
Not Bashful	1.3	Takes Chances	6.0	Takes Chances	5.6
Takes Chances	.8	Good at Games	2.2	Good at Games	3.3
		Not Bashful	1.2	Not Bashful	2.7
		Leader	1.1	Can Take Joke	1.4
				Leader	.8
				Full of Fun	.1

boys' means. Secondarily, there is a tendency for the girls' means to shift in the unfavorable direction, at least on certain traits. Dimensions which show most clearly the upward trend with grade level in boys' mean scores and the reverse in girls' are POPULAR, FULL OF FUN, NOT QUARRELSOME, NOT BOSSY, GOOD AT GAMES, and, most conspicuously, REAL BOY and ACTS LIKE LITTLE LADY. Other item-pairs of the complete test might also have exhibited this phenomenon, but since they were added late in the testing program, comparative data for various grade levels are not available.

The age trend noted above is strikingly consistent for the item-pairs mentioned, but the grade-level differences are usually smaller than the sex differences and are rarely statistically significant. Indeed, in the case of girls' means, no dimensions show a significant increase in score with advancing grade level, but some of those which show a marked drop from Grade I to Grade III show minor gains from Grade III to Grade V.

D. SUMMARY OF SEX AND GRADE DIFFERENCES IN MEAN SCORE BY ITEM-PAIR

1. *Dimensions Included at Grade Levels I, III, and V*

1–2 WIGGLY–QUIET. The average girl is rated quiet by approximately six per cent of her classmates, at all grade levels, while the average boy is considered restless by about four per cent. Differences in mean score between grade levels are not significant for either sex, but the critical ratios of the difference between girls and boys range from 4.0 to 8.0.

4–3 NOT MANY FRIENDS–POPULAR. The average girl is "liked by everyone" at all levels, but decreasingly so, while the score of the average boy, which is slightly below the zero point in the first and third grades, rises in the fifth grade to a point which is positive although lower than that for girls. Sex differences in favor of girls are significant at Grades I and III but not at Grade V. Grade differences are not significant for girls, but for boys the improvement from third to fifth grades reaches significance (C.R. = 3.0).

6–5 SERIOUS–FULL OF FUN. The average girl is judged to be FULL OF FUN, though her score is smaller than on POPULAR and QUIET. This item is one of the few on which boys enjoy favorable status at all grade levels, but though their mean score increases with age, it does not overtake a diminishing score for girls until the fifth grade. Sex differences are not significant, nor are age differences, although the drop in the girls' mean from Grade I to Grade III and the gain in the boys' mean from Grade III to Grade V approach the conventional level (C.R.'s = 2.1 and 2.4, respectively).

7–8 QUARRELSOME–NOT QUARRELSOME. At all levels girls are rated NOT QUARRELSOME and boys at the opposite side of the scale. This difference is statistically significant at all grade levels but becomes less marked with age, and at the fifth-grade level the average boy is regarded as QUARRELSOME by only one per cent of the class. Grade differences are not statistically significant save for the boys' gain in score from Grade III to Grade V (C.R. = 3.1).

9–10 DOESN'T TAKE CHANCES–TAKES CHANCES. This is a unique item-pair, in that it is the only one on which boys enjoy more favorable status than girls at all levels, and the only one on which the average girl received a negative score (Grades III and V only). Moreover, it shares with GOOD AT GAMES, NOT A SHOW-OFF, and NOT BASHFUL the distinction of reversing the usual tendency for boys' and girls' scores to approximate each other more closely with increasing age. The sex difference is significant in Grades III and V. Age differences are negligible for boys, but the drop in the average girls' score between Grades III and V approaches the level of significance. Thus the trend seems to be more dependent upon the girls growing timid than upon the boys growing bolder. Apparently the older girls learn to conform to the feminine pattern of our society and suppress their first-grade venturesomeness.

11–12 BOSSY–NOT BOSSY. NOT BOSSY is applied to the girls at all levels, its converse applied to the boys. The alleged bossiness of boys declines with age, and the fifth-grade mean scores for both sexes come to lie very near the zero point. However, girls are rated significantly less bossy than boys in Grades I and III (C.R.'s = 2.8 and 4.4, respectively). Age differences are not significant on this dimension.

13–14 POOR SPORT–GOOD SPORT. This item-pair conforms to the pattern of other results, if not to the preconceptions of the writer, in that girls are consistently rated better sports than boys. However, both groups are scored on the positive side, and the trend leads us to make the extrapolation that by the seventh grade, perhaps, boys would be considered at least as good sports as the girls. Sex and age differences are not statistically significant on this dimension.

15–16 BASHFUL–NOT BASHFUL. NOT BASHFUL shows no consistent grade trends either for girls or for boys. Both sexes are at all levels rated on the positive side with boys receiving a larger mean score than girls. The boys' mean score increases between the third and fifth grades to a degree which approaches significance (C.R. = 2.0) and is probably significantly higher in Grade V than the girls' (C.R. = 2.7).

18–17 Not Good at Games–Good at Games. This item-pair resembles the Takes Chances dimension, in that boys' scores grow more favorable with age while girls' scores decline. However, both the average boy and the average girl receive a favorable score at all grade levels. Age differences for both sexes are somewhat below the conventional significance point. (C.R.'s fluctuate around 1.0 for boys and 2.0 for girls.) However, sex differences are by the fifth grade definitely reliable (C.R. = 3.3). The judges seem to regard this dimension as more or less the exclusive province of boys, and mentions of girls at either pole become relatively infrequent among older children.

19–20 Gets Mad Easily–Doesn't Get Mad Easily. Scores on this dimension are not extreme, the girls remaining on the positive side, the boys on the negative at all levels. Age differences are not significant, but boys are considered more prone to get mad, especially at the third-grade level (C.R. = 5.5).

21–22 Sissy–Real Boy; **23–24** Tomboy–Acts Like Little Lady. These item-pairs correspond for boys and for girls and may conveniently be considered together. Here again the usual trend for boys' scores to increase and for girls' to decrease is extremely marked and very consistent, although the mean for both sexes is toward the positive pole at all levels. The critical ratio of the difference in mean score at adjacent age levels is around 2.0 for boys and 1.0 for girls. It will be recalled that the wording both of the favorable items and their opposites was changed for fifth-grade classes, since it was felt that the item Sissy, as originally phrased, was too emotion-loaded, while the comments of children led us to believe that the Little Lady item-pair was tapping attitudes about neatness and fussiness rather than about femininity.[1] However, the trend of the curves is perfectly regular and apparently unaffected by the revisions, the boys becoming more boyish and the girls less "ladylike" with increasing age.

25 Best Friend. This item, which has no antonym, shows slightly increasing scores with age, reflecting heavier voting on the question. Sex differences are negligible, although girls have slightly higher scores at all three age levels. The average child is mentioned by from 5 per cent to 6 per cent of his classmates as a Best Friend, the equivalent, in an average class of 30, of slightly more than 1½ votes received per child.

2. *Dimensions Included Only at Grade Levels III and V*

27–26 Not Good-Looking–Good-Looking. As might be expected, the girls receive favorable scores on Good Looking, while the boys are rated slightly below the indifference point. Age differences on this item are negligible, but the sex difference is marked at both grade levels for which the item was included (C.R.'s = 3.8 and 2.6 for Grades III and V, respectively).

28–29 Show-Off–Not a Show-Off. This is the only item-pair on which boys' scores grow less favorable with age. They are rated definite show-offs in the third grade and grow more extreme in the fifth grade. On the other hand, the girls enjoy a favorable status at both levels which increases with age. Sex differences are significant at both the third- and fifth-grade levels (C.R.'s = 3.2 and 5.7, respectively). The difference in mean score between adjacent grade levels is only about 1.5 times its standard error for both boys and girls.

---

[1] See Chapter II, Table 2.

31–30 NOT A LEADER–LEADER. LEADER is an attribute on which both the average boy and average girl receive favorable though small scores. The mean for boys slightly exceeds that for girls. For some undetermined reason, the scores for both sexes decrease in magnitude with age, though not to a significant degree.

3. *Dimensions Included Only at Grade Level V*

33–32 UNTIDY–TIDY. TIDY, an item for which we possess only fifth-grade data, seems to be an exclusive attribute of girls. The average girl is so regarded by eight per cent of her class, the average boy being named as untidy by one per cent. The critical ratio of this difference (C.R. = 7.9) is larger than that on any other dimension at this grade level save WIGGLY–QUIET.

34–35 FIGHTS–DOESN'T FIGHT. This is a dimension on which girls receive a favorable, boys an unfavorable score. The sex difference in means is four times its standard error.

37–36 NOT FRIENDLY–FRIENDLY. Both boys and girls are considered friendly, the girls to a significantly greater degree than the boys (C.R. = 4.0).

39–38 CAN'T TAKE A JOKE–TAKES A JOKE ON SELF. Both sexes are rated as able to take jokes on themselves, but here boys lead the girls by an insignificant margin.

E. VARIABILITY IN SCORES

Standard deviations of score distributions for sex-grade groups have been presented in Table 4. Sex and grade comparisons tend to parallel those for the means cited above. As may be seen by inspection of the table, there is greater variability in boys' than in girls' scores at all three grade levels, probably because judges of both sexes are more given to naming boys than girls for undesirable qualities. On the other hand, score variability shows little consistent relation to grade level, the average standard deviation dropping very slightly from Grade I to Grade III and increasing fairly markedly from Grade III to Grade V. The decline in variability from Grade I to Grade III is possibly the effect of sampling error in the first-grade results, since the total population was small. To account for the increase from Grade III to Grade V, two other factors may be suggested. First, the tendency of older judges to name more people per item results in the identification of a larger proportion of the class. Second, the increased agreement found among older judges operates to reduce the number of zero scores produced by cancellation, and thus increases the standard deviations.

When individual dimensions are considered, it is obvious that several different factors contribute to differences in score variability. The inclusion of WIGGLY–QUIET and NOT MANY FRIENDS–POPULAR among those having large sigmas may result in part from the fact that they occur at the beginning of the test where voting is heaviest. The low variability of CAN'T TAKE JOKE–TAKES JOKE ON SELF is to be anticipated from its position at the end of the test, where mentions are few. On the other hand, SISSY–REAL BOY and TOMBOY–ACTS LIKE LITTLE LADY, which occur well along in the test, have very large standard deviations. Items having large variabilities for one sex only, such as GOOD-LOOKING for girls and LEADER and GOOD AT GAMES for boys, seem to have a special relevance for children of the one sex that they do not possess for the opposite. In general it seems safe to say that the item-pairs yielding large variabilities in scores do so not only because they are more easily rated, but also because they refer to qualities of personality which seem important to the children themselves.

F. THE PROPORTIONATE DETERMINATION OF SCORES BY JUDGES OF THE SAME AND OF THE OPPOSITE SEX

In considering the tendency for girls to enjoy on the average more favorable reputations than boys, one may raise the question as to the degree to which the reputations of boys and girls are determined by members of the same and of the opposite sex. Figure 2 presents our findings on this point. As may be gathered by inspection, on most items somewhat more than half of the votes received are cast by judges of the same sex. This tendency toward determination of scores by judges of the same sex seems to increase slightly with grade level on most items.

Scores on a few item-pairs, however, are determined for both sexes by about equal numbers of votes from boys and from girls at all grade levels studied. Dimensions of this sort include SISSY–REAL BOY, TOMBOY–ACTS LIKE LITTLE LADY, NOT GOOD-LOOKING–GOOD-LOOKING, and WIGGLY–QUIET. Since judges were urged to vote on as many items as possible and since the REAL BOY and LITTLE LADY dimensions refer to persons of one sex only, it is to be expected that scores on these variables are determined by boys and girls in proportion to their frequency in the total population of judges. In the case of at least the WIGGLY–QUIET pair, data on judgments to be presented in Chapter IV show that the equal contribution of boy and girl judges in determining scores was produced by a tendency of girls to mention boys as wiggly and of boys to mention girls as quiet.

On every dimension included in the first-grade test, girls' scores were determined to a greater degree by votes from boys than were boys' scores determined by votes from girls. In the fifth grade, on every dimension save BEST FRIEND, the opposite situation obtained—boys being mentioned by girls relatively more frequently than girls were mentioned by boys. Results for the third grade fluctuate from item to item between these opposite tendencies. All these differences in relative determination of scores by judges of the same and opposite sex are, however, quite small and derive their significance from the consistency of the trend on all dimensions, rather than from the numerical values of the percentages.

In the first grade the finding is in part an artifact, since boys constituted 59 per cent of the total group of judges and should be expected on a chance basis to determine by an equivalent percentage the scores received by girls. However, in the fifth grade the situation had reversed itself, though the boys were still slightly more numerous than girls in the population studied.

Our data are only suggestive, but the findings cited are at least congruent with the hypothesis that age for age, girls are more mature than boys socially as well as physiologically, the girls taking the lead both in sex segregation of play groups at the primary grade level, and

PERCENTAGE DETERMINATION BY SAME SEX AND BY OPPOSITE SEX OF BOYS' AND OF GIRLS' SCORES ON THE REPUTATION TEST

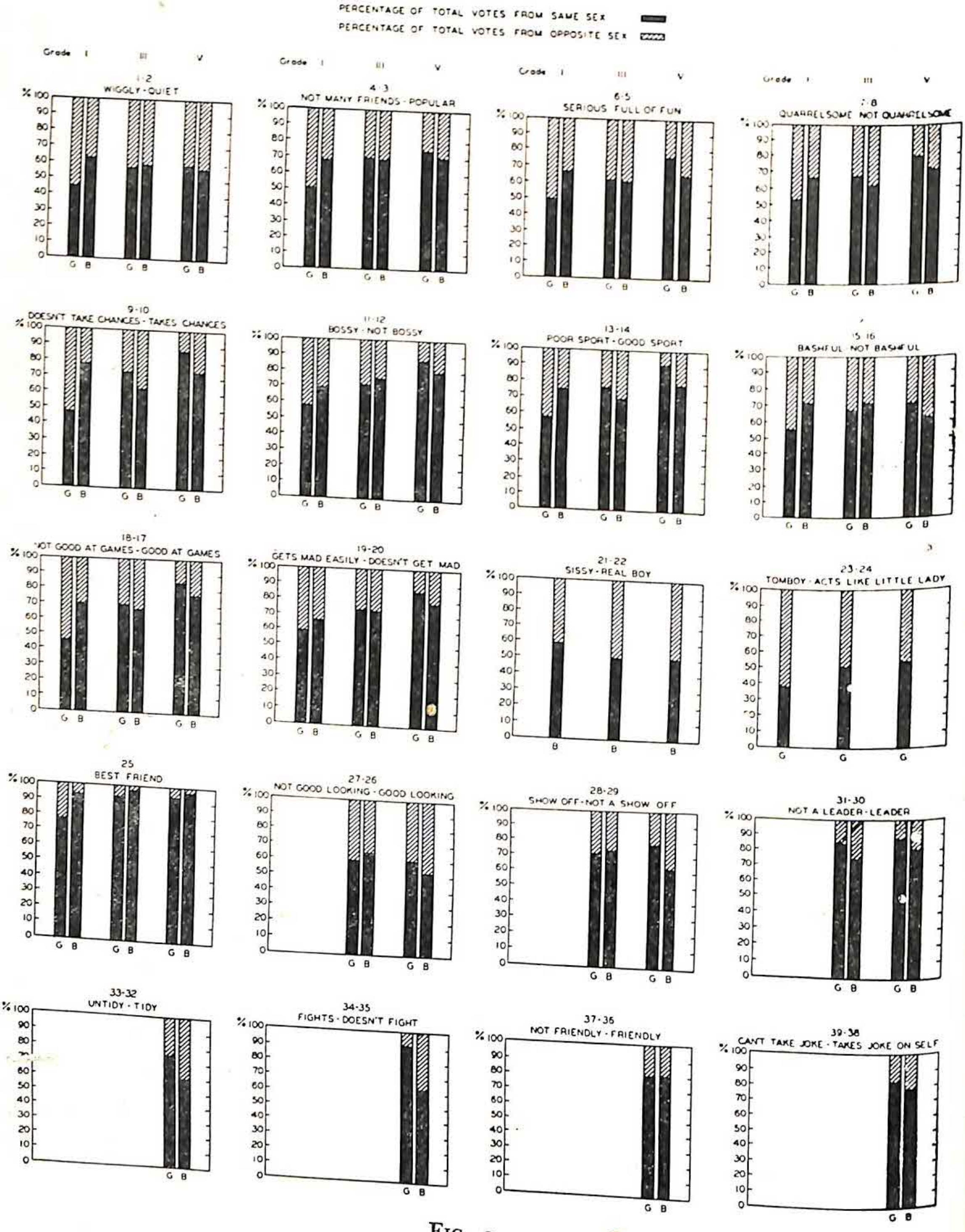


FIG. 2

in heterosexual interest at the beginning of adolescence.

G. SUMMARY OF CHAPTER III

Scores for children on each item-pair were derived by algebraic summation of favorable and unfavorable mentions and corrected for differences in the size of the classrooms. Distributions of these scores were found to show a considerable piling up of zero scores. However, the proportion of children receiving a definite characterization on one pole or the other of a dimension increased with grade level, and on the average item-pair only one child in five received a score of zero.

Children's attitudes toward each other, insofar as they can be inferred from the mean score of girls and of boys on each dimension, seem to have been anticipated quite accurately and succinctly by the famous nursery rhyme about little girls being made from sugar and spice, and little boys from snakes and snails. More specifically, the typical girl is judged to be QUIET, POPULAR, FULL OF FUN, NOT QUARRELSOME, A GOOD SPORT, A LITTLE LADY, GOOD-LOOKING, NOT A SHOW-OFF, TIDY, and FRIENDLY. The typical boy is WIGGLY, QUARRELSOME, BOSSY, and a SHOW-OFF. On the favorable side of the ledger he is inclined to TAKE CHANCES, and be NOT BASHFUL, GOOD AT GAMES, and a REAL BOY. On most of the remaining qualities covered by the test, girls enjoy more favorable status than boys, though to a less extreme degree.

Age differences in mean scores tend to be less significant than the sex differences cited above. However, a trend with increasing grade level for girls to lose status and for boys to improve, operates to reduce the magnitude of the sex differences. No hypothesis is advanced to explain the drop in girls' scores, but the improvement noted for boys is probably in part the result of the reluctance of the older judges to make characterizations of each other on the unfavorable items.

Both boys and girls receive their scores primarily from individuals of their own sex, and this tendency grows slightly more marked with age. However, girls' scores are in the first grade determined by boy judges to a relatively greater degree than are boys' scores determined by girls. This trend disappears around the third-grade level, and is replaced in the fifth grade by a tendency for girls to contribute more heavily to boys' scores than vice versa.

At all three grade levels studied, votes on the item BEST FRIEND are received almost exclusively from children of the same sex as the person named.

The picture of the typical boy and girl as conceived by children even in the primary grades seems to be almost a photostat of the common identification by adults in our society of aggressiveness, restlessness, and daring with masculinity, and of amiability, docility, and timidity with femininity. Perhaps equally significant for the theory of adult personality organization is the fact that especially in the lower grades the traits associated with girls tend to be positively valued, while many of the traits associated with boys are disapproved, a situation which may have important implications for the attitudes of adult men and women toward themselves and toward each other.

CHAPTER IV

PATTERNS IN JUDGMENT

THE Reputation Test yields data of two sorts about each child: first, the judgments others make about him, i.e., his "reputation," and second, the judgments he makes. Normative findings of the first sort have been considered in the previous chapter. In the present one, the frame of reference will be inverted and findings presented on the characteristics of boys and girls of various grade levels *as judges of others and of themselves.* Data on teachers' judgments of their classes will also be included here, because they provide a contrast to the children's judgments both as to items avoided or emphasized in voting, and as to the identification of certain attributes with masculinity and femininity. Some of the results have been anticipated in the last chapter, owing to the reciprocality between group patterns in votes received and in votes given. However, in the present chapter, analyses of judgments will be based on individual items rather than item-pairs, and patterns concealed or obscured by treatment in terms of item-pairs will emerge more clearly.

Several purposes are served by this analysis. First, results may be used as a guide to the selection and phrasing of test items, the choice of optimal test length, etc. Further, the investigation of judgment patterns contributes to an understanding of childhood socialization by revealing in marked emphases or avoidances of judgment on particular items, the group values and group pressures to which the individual must adjust. Lastly, it is necessary to secure normative data for the judgments (and projections) of the group in order to evaluate individual children's voting records as projections of their own personalities.

A. CHILDREN'S JUDGMENTS OF EACH OTHER

In Table 6 are given the mean number of votes cast by boy and girl judges for members of their own and of the opposite sex. Since judgments are demanded on individual items, the data have likewise been so analyzed, and favorable and unfavorable traits have been segregated in this table. The original order of presentation of the questions is indicated by the item number, save in the case of BEST FRIEND, which was always given last. Self-mentions are not included in this table, nor is any correction made for variation in the size of the classrooms, since the number of others mentioned by a child on a given item was hardly ever more than a fraction of even the smallest class. Although instructions were phrased so as to permit the children to name as many others on each question as they felt merited the description, there were few traits on which the average child mentioned as many as three persons. The modal number of votes per child on each item was one in most cases. However, the distributions are all noticeably skewed since it was not uncommon for a child to mention as many as eight or ten others on a few items which had special significance for him.

1. *Grade Differences in Judgments*

As may be seen from Table 6, the mean number of votes cast per child on any given item does not vary greatly from grade to grade; nor is there any consistent trend, the average for all items dropping from Grade I to Grade III, and increasing from Grade III to Grade

TABLE 6

Mean Number of Votes Cast per Judge on Individual Items of the Reputation Test

Item	Grade I				Grade III				Grade V			
	Girl Judges ($N=41$)		Boy Judges ($N=65$)		Girl Judges ($N=318$, items 1–25) ($N=169$, items 26–31)		Boy Judges ($N=336$, items 1–25) ($N=184$, items 26–31)		Girl Judges ($N=296$)		Boy Judges ($N=312$)	
	Votes For Girls	Votes For Boys	Votes For Girls	Votes For Boys	Votes For Girls	Votes For Boys	Votes For Girls	Votes For Boys	Votes For Girls	Votes For Boys	Votes For Girls	Votes For Boys
Unfavorable Items												
1 Wiggly	.3	2.2	.4	1.7	.4	1.1	.3	1.2	.6	1.4	.4	1.4
4 Not Many Friends	.4	1.3	.4	1.2	.6	.8	.3	1.0	1.3	.8	.5	1.0
6 Serious	.6	1.0	4	1.1	.6	.5	.4	.7	1.2	.6	.5	.9
7 Quarrelsome	.3	1.7	.2	1.8	.5	1.1	.2	1.4	1.2	.8	.3	1.5
9 Doesn't Take Chances	.5	.5	.6	1.0	.5	.3	.4	.5	1.5	.3	.3	1.3
11 Bossy	.6	1.2	.3	1.2	1.0	.4	.3	1.1	1.7	.6	.2	1.6
13 Poor Sport	.5	1.0	.4	1.2	.7	.8	.2	1.2	1.2	.5	.1	1.4
15 Bashful	.7	.7	.6	.8	.6	.2	.4	.5	1.5	.4	.6	.6
18 Not Good at Games	.4	1.0	.4	1.2	.7	.6	.4	1.0	1.6	.4	.4	1.6
19 Gets Mad Easily	.7	1.2	.4	1.4	.9	.6	.2	1.3	1.5	.5	.2	1.5
21 Sissy	—	1.7	—	1.3	—	1.3	—	1.0	—	1.4	—	1.3
23 Tomboy	1.0	—	1.0	—	.8	—	.9	—	1.8	—	1.5	—
27 Not Good-Looking					.6	.4	.4	.5	1.3	.8	.7	.8
28 Show-Off					.8	.5	.3	1.1	1.1	1.3	.3	1.4
31 Not a Leader					.5	.3	.1	.7	1.4	.3	.2	1.2
33 Untidy									.7	1.0	.1	1.1
34 Fights									.6	1.2	.0	1.4
37 Not Friendly									1.0	.3	.2	.8
39 Can't Take a Joke									.9	.2	.2	.8
Avg. Votes per Item per J. (Unfav.)	.5	1.2	.5	1.3	.6	.6	.3	.9	1.2	.7	.4	1.2
Favorable Items												
2 Quiet	1.7	.8	1.1	1.5	1.3	.3	1.0	.7	2.0	.4	1.3	.9
3 Popular	1.8	.7	.9	1.4	1.8	.3	.6	1.3	2.6	.6	.7	2.2
5 Full of Fun	1.5	1.0	1.0	1.3	1.1	.6	.6	1.0	2.2	.8	.5	1.7
8 Not Quarrelsome	1.5	.6	.8	1.0	1.3	.3	.6	.8	2.3	.5	.6	1.6
10 Takes chances	1.0	.7	.5	1.7	.8	.8	.1	1.3	1.2	1.0	.1	2.0
12 Not Bossy	1.0	.4	.4	1.1	1.2	.2	.5	.8	1.8	.3	.3	1.6
14 Good Sport	1.5	.6	.6	1.8	1.6	.6	.4	1.7	2.3	.5	.2	2.2
16 Not Bashful	1.2	.7	.4	1.3	1.1	.4	.4	1.0	1.7	.9	.6	1.5
17 Good at Games	1.2	.8	.8	1.4	1.2	.8	.3	1.8	2.0	1.0	.3	2.9
20 Doesn't Get Mad Easily	1.2	.8	.4	1.1	1.4	.3	.6	1.0	2.0	.4	.4	1.6
22 Real Boy	—	1.7	—	1.9	—	1.8	—	2.0	—	2.9	—	2.8
24 Acts Like Little Lady	1.8	—	1.9	—	1.7	—	1.4	—	2.4	—	1.8	—
25 Best Friend	1.2	.1	.2	1.2	1.5	.0	.1	1.4	1.9	.1	.1	1.8
26 Good-Looking					1.3	.2	.7	.5	1.7	.7	1.2	.7
29 Not a Show-Off					1.1	.2	.4	.7	1.8	.4	.5	1.2
30 Leader					.8	.4	.1	1.2	1.5	.4	.2	1.6
32 Tidy									2.6	.5	.9	1.2
35 Doesn't Fight									1.4	.5	.2	1.3
36 Friendly									2.2	.2	.6	1.4
38 Takes a Joke on Self									1.3	.4	.2	1.4
Avg. Votes per Item per J. (Favor.)	1.4	.7	.8	1.4	1.3	.5	.5	1.1	1.9	.7	.6	1.7
Avg. Votes per Item per J. (U. & F.)	1.0	1.0	.6	1.3	1.0	.6	.4	1.0	1.6	.7	.5	1.4

V. The Grade I results are based on relatively few cases and are hence less accurately determined. The increase in mean number of votes per item between Grades III and V may be as much the result of a change in the mode of test administration, as of a true age difference in the subjects' voting habits. (It will be recalled that fifth graders voted more freely on uncomplimentary items when they wrote them down, than when they verbalized them to an interviewer.)

12.8.70
9157

Although age differences in volume of voting over the test as a whole are not marked, there is considerable variation among items at all grade levels. Inspection of the data shows that the most-used items tended to occur near the beginning of the test, while the least-used items occurred near the end where the children were hampered by fatigue or lack of time. At all grade levels the last item in the test was in the least-mentioned group. (In the fifth grade the item BEST FRIEND actually came last in the booklet, but children were requested to fill it in five minutes before the end of the period and then go back and work on the others.)

A second factor, perhaps even more important than fatigue in limiting the volume of voting, was the tendency of the child judges to avoid making uncomplimentary nominations. At all three grade levels the four most often mentioned traits were favorable, and of the four least often mentioned traits, all but BEST FRIEND (Grade I only) were unfavorable. This tendency grew more pronounced with older children. Taking all dimensions together, the average critical ratio of the difference between the mean number of votes per positive item and the mean number of votes per corresponding negative item was 1.97 in the first grade in favor of the positive pole. In the third grade the same statistic increased to 5.39 and in the fifth grade to 7.09. Only a very few negative items, BOSSY, GETS MAD, QUARRELSOME, and SHOW-OFF, elicited more voting than their positive antonyms and in such cases none of the critical ratios of the difference between negative and positive means was statistically significant. Apparently, the children reversed their tendency to avoid mentioning others as having undesirable traits only with respect to a few relatively innocuous items. For example, the items SHOW-OFF and BOSSY are correlated with LEADER and TAKES CHANCES and hence carry much less opprobrium than items such as SISSY and NOT MANY FRIENDS. It is unlikely that undesirable attributes are actually as much rarer or as much more difficult to observe as these findings would imply, especially since in many cases the undesirable item refers to more specific behavior than the desirable one, and since only the *relatively* more extreme children were to be cited at either end of the scale; but the socializing process to which children in our culture are subjected seems to have led them even in the first grade to regard the making of a formal statement uncomplimentary to someone else as a violation of the mores.

2. *Sex Differences in Judgments*

Considering the test as a whole, there is a slight though consistent tendency for girls to cast more votes than boys at all three grade levels studied, a trend that increases with grade level. In the first and third grades the trend is small, but by the fifth grade, the average critical ratio of the difference between the mean number of votes by girl and by boy judges for all items together is 3.0, and on many individual items the critical ratios are much larger. At no grade level was there an item which elicited a significant difference in volume of voting in favor of the boys. The Reputation Test as used in this investigation contains numerous items such as FIGHTS, GOOD AT GAMES, TAKES CHANCES, etc., which it had been felt might overload the test with questions about behavior which may be more characteristic of boys and possibly more easily observed by them. However, so far as vote casting is concerned, the girls clearly lead the boys, a finding which has been reported by Campbell (1) and Tryon (10) for an older age group. Possible explanations are that a difference in motivation to take the test exists, that girls are more interested than boys in social relationships, or that girls are able to verbalize their observations of others more readily.

Further questions of interest are whether girl and boy judges differ in the degree to which they express opinions about members of the opposite sex, whether boy and girl judges tend to give more favorable mention to members of their own or to members of the opposite sex, and whether there are certain items on which judges of both sexes vote primarily for one sex to the exclusion of the other.

The first striking fact to be observed is that, in general, children vote much more heavily for persons of their own than of the opposite sex. This finding has been noted for adolescents by Campbell, and has been found in the present investigation to hold true for both positive and negative aspects of reputation, save in the case of first-grade girls, who voted more heavily for boys on unfavorable items. However, there exists an amusing difference between the voting habits of girl and boy judges. At all three grade levels, boys cast more favorable than unfavorable votes, both for their own and for the opposite sex. Girls cast more favorable than unfavorable votes for their own sex, but more unfavorable than favorable votes for the opposite sex. Table 7 shows this clearly. Especially in the first grade, but to a lesser degree throughout the elementary years, the unwillingness of children to make negative judgments about each other does not extend to girls' opinions about boys.

By inspection of Table 6, it is possible to locate items which both boy and girl judges consider to be more characteristic of boys and other items which are more frequently applied to girls. In Grade I, both boys and girls voted more heavily for boys on all the negative items. In Grade III, both sexes mentioned more boys on WIGGLY, QUARRELSOME, POOR SPORT, and NOT MANY

TABLE 7
AVERAGE VOTES PER ITEM BY BOY AND GIRL JUDGES FOR SAME AND OPPOSITE SEX

Type of Vote	Grade I		Grade III		Grade V	
	Favorable	Unfavorable	Favorable	Unfavorable	Favorable	Unfavorable
Boy Judges:						
Average votes per item for boys	1.4	1.3	1.1	.9	1.7	1.2
Average votes per item for girls	.8	.5	.5	.3	.6	.4
Girl Judges:						
Average votes per item for girls	1.4	.5	1.3	.6	1.9	1.2
Average votes per item for boys	.7	1.2	.5	.6	.7	.7

FRIENDS, and more girls on GOOD-LOOKING, QUIET, and NOT A SHOW-OFF. In Grade V, both sexes mentioned more boys on WIGGLY, SHOW-OFF, UNTIDY, and FIGHTS, and more girls on GOOD-LOOKING and QUIET. Additional items on which boys voted for boys to a greater degree than girls voted for girls are TAKES CHANCES, GOOD AT GAMES, and LEADER. The consistency with which traditionally masculine and feminine attributes are so identified in the primary grades is impressive.

B. CHILDREN'S SELF-EVALUATIONS

In Table 8 are presented data on the percentage frequencies with which boys and girls mentioned themselves on the various favorable and unfavorable items on the Reputation Test. The item BEST FRIEND is omitted from this table, although in the course of the entire testing program several either unusually frank or unusually flippant children so named themselves. In no case did a child mention himself on opposite poles of the same dimension.

The most striking feature of Table 8 is the very much greater frequency of self-mentions on positive items than on negative ones. This finding holds true both for boys and for girls and at all three grade levels studied. On only a few negative items did children name themselves more often than on the corresponding positive traits, and such instances were in most cases relatively innocuous. The unfavorable items on which *girls* named themselves more often than on the paired favorable traits are: WIGGLY (Grade III), and BASHFUL, TOMBOY, and NOT GOOD-LOOKING (Grade V). Such items for boys are: WIGGLY and NOT GOOD-LOOKING (Grades III and V).

Although positive self-mentions were more frequent than negative ones at all levels, the number of children making self-mentions on the average *favorable* item *decreased* from around 28 per cent in Grade I to around 22 per cent in Grade V, while the number making *unfavorable* self-nominations *increased* over the same range approximately from 3 per cent to 9 per cent. This trend obtained among children of both sexes. On many individual items, the difference in frequency of self-nomination at adjacent age levels in accordance with this trend is highly significant (C.R.'s ≥ 3.0).

In comparing self-evaluations with similar tendencies in judging others, one notes that judgments of both types are more frequent on favorable than on unfavorable items. However, the tendency of favorable self-judgments to decline with grade level while unfavorable self-judgments increase, stands in striking contrast to the children's age trend in judging others to avoid more and more the seriously derogatory mentions in favor of the complimentary ones.

> Social learning may well be responsible for both situations. In judging others, there seems to arise in older children an unwillingness to express criticism or hostility too frankly. In judging themselves, the children appear not only to have grown more capable of detachment, but also to have assimilated at least in part the taboo which society places upon too obvious self-congratulation. Nevertheless, it is probably inaccurate to assume that the younger children who named themselves for favorable traits always felt that they really possessed them, or that the decrease in such mentions necessarily reflects an increase in modesty. Analysis of the records of individuals indicates that probably many complimentary self-mentions of first graders were made with reference to areas in which they felt insecure and were moved to "protest too much." Older children came to avoid making favorable self-mentions on the dimensions on which they feared that class opinion would be discrepant, but only rarely named themselves for qualities seriously disapproved by the group.

Referring again to Table 8, one may

TABLE 8

PERCENTAGE FREQUENCY OF CHILDREN'S MENTIONS OF THEMSELVES ON INDIVIDUAL ITEMS OF THE REPUTATION TEST

Item	Grade I		Grade III		Grade V	
	Girls ($N=41$)	Boys ($N=65$)	Girls ($N=318$, items 1–25) ($N=169$, items 26–31)	Boys ($N=336$, items 1–25) ($N=184$, items 26–31)	Girls ($N=296$)	Boys ($N=312$)
Unfavorable Items						
1 WIGGLY	5%	5%	11%	18%	5%	14%
4 NOT MANY FRIENDS	2	2	4	7	7	6
6 SERIOUS	5	6	4	6	5	5
7 QUARRELSOME	0	2	6	8	6	8
9 DOESN'T TAKE CHANCES	5	6	6	4	19	16
11 BOSSY	0	5	2	2	4	7
13 POOR SPORT	0	3	2	4	4	5
15 BASHFUL	5	8	18	15	35	17
18 NOT GOOD AT GAMES	5	5	9	11	15	13
19 GETS MAD EASILY	2	5	4	9	11	10
21 SISSY		0		2		3
23 TOMBOY	2		18		38	
27 NOT GOOD-LOOKING			7	7	10	8
28 SHOW-OFF			1	8	2	8
31 NOT A LEADER			5	11	8	7
33 UNTIDY					2	6
34 FIGHTS					4	10
37 NOT FRIENDLY					0	3
39 CAN'T TAKE A JOKE					1	2
Mean (Unfavorable Items)	2.8%	4.3%	6.9%	8.0%	9.7%	8.2%
Favorable Items						
2 QUIET	22%	29%	8%	11%	10%	6%
3 POPULAR	20	15	21	22	8	11
5 FULL OF FUN	15	22	33	35	27	21
8 NOT QUARRELSOME	37	32	48	49	24	23
10 TAKES CHANCES	22	45	21	44	20	30
12 NOT BOSSY	22	14	23	35	23	21
14 GOOD SPORT	27	37	27	44	23	30
16 NOT BASHFUL	24	28	24	33	20	26
17 GOOD AT GAMES	24	43	33	53	22	34
20 DOESN'T GET MAD EASILY	29	37	34	33	19	20
22 REAL BOY		51		55		46
24 ACTS LIKE LITTLE LADY	24		28		20	
26 GOOD-LOOKING			7	3	8	6
29 NOT A SHOW-OFF			17	17	17	13
30 LEADER			14	30	23	28
32 TIDY					28	19
35 DOESN'T FIGHT					32	33
36 FRIENDLY					17	19
38 TAKES A JOKE ON SELF					38	35
Mean (Favorable Items)	24.2%	32.1%	24.1%	33.1%	21.1%	23.4%
Mean (All Items)	13.5%	18.2%	15.5%	20.6%	15.4%	15.8%

observe a fairly constant tendency for boys to vote for themselves more frequently than do girls, whereas in voting for others, the girls made more nominations. The average differences are small but are consistent save in the case of unfavorable items at the fifth-grade level, on the average of which 9.7 per cent of the girls mentioned themselves and only 8.2 per cent of the boys. On individual items referring to both sexes, there are in Grade I no statistically significant sex differences in the percentage frequency of self-mentions, but in Grades III and V the boys lead by a significant margin on several items, including GOOD AT GAMES, WIGGLY, SHOW-OFF, and FIGHTS. On no item save BASHFUL (Grade V only) do girls make significantly more self-nominations than boys.

Whether boys are genuinely as much more egocentric as these results imply is a debatable question. However, the fact that the differences between boys' and girls' voting, both for themselves and for others, parallel the differences between first-grade and fifth-grade patterns for the combined group of judges of both sexes, suggests that girls are exhibiting a somewhat more mature type of evaluation.

C. TEACHERS' JUDGMENTS OF CHILDREN

It will be recalled that in order to secure a measure of adult opinion about children expressed in the same form as the children's opinions of each other, the interviewers gave to the teacher of each classroom the same Reputation Test taken by her pupils. Unfortunately, the number of teachers included was too small to render their statements a statistically adequate basis for generalization. Further, in this grade range, each class had but one teacher for all subjects. Since all the teachers were women, it is impossible to know to what degree their opinions would have coincided with those of an equal group of male adult observers. Nevertheless, this body of data, when compared with that supplied by the children themselves, reveals interesting and suggestive differences between adult and juvenile points of view.

In Table 9 are presented the results of the voting by teachers. Findings are given here in terms of the percentage of boys and of girls mentioned by their teachers on each item.

Like the children, the teachers voted more frequently on complimentary than on uncomplimentary items. This finding is true of mentions of both boys and girls and at all three grade levels studied, but the actual frequencies of teachers' votes are so low that the difference is probably not significant. On a few dimensions the teachers voted slightly more heavily on the unfavorable trait than on its favorable antonym. Among these are WIGGLY (for boys, all grades), QUARRELSOME (for boys, Grades III and V; for girls, Grade I), GETS MAD EASILY (for boys, Grades III and V), FIGHTS (for boys, Grades III and V), and DOESN'T TAKE CHANCES (for girls, Grades III and V). Save for the last, these items all refer to behavior which disrupts classroom routine, and to which teachers are in consequence sensitized. It will be recalled that among the few unfavorable items upon which the children mentioned others more frequently than on the corresponding desirable traits, BOSSY was the only one appearing at all grade levels, and in upper grades the items GETS MAD and QUARRELSOME dropped out in favor of SHOW-OFF. The item WIGGLY, which elicited so many votes from teachers, did not appear at all in

TABLE 9

PERCENTAGE FREQUENCY OF MENTIONS OF CHILDREN BY TEACHERS ON INDIVIDUAL ITEMS OF THE REPUTATION TEST

Items	Grade I		Grade III		Grade V	
	Girls ($N=49$)	Boys ($N=70$)	Girls ($N=340$, items 1–25) ($N=178$, items 26–31)	Boys ($N=352$, items 1–25) ($N=194$, items 26–31)	Girls ($N=306$)	Boys ($N=322$)
Unfavorable Items						
1 WIGGLY	8%	16%	7%	13%	4%	13%
4 NOT MANY FRIENDS	4	10	4	6	5	6
6 SERIOUS	10	9	7	6	7	5
7 QUARRELSOME	12	6	6	11	5	11
9 DOESN'T TAKE CHANCES	10	7	5	3	8	2
11 BOSSY	10	9	9	6	6	7
13 POOR SPORT	0	4	2	9	1	7
15 BASHFUL	12	10	7	7	9	6
18 NOT GOOD AT GAMES	10	7	5	6	6	6
19 GETS MAD EASILY	6	9	3	9	5	8
21 SISSY		11		5		8
23 TOMBOY	6		7		7	
27 NOT GOOD-LOOKING			3	3	4	1
28 SHOW-OFF			6	9	2	9
31 NOT A LEADER			6	8	7	7
33 UNTIDY					1	7
34 FIGHTS					3	7
37 NOT FRIENDLY					2	4
39 CAN'T TAKE A JOKE					3	4
Mean (Unfavorable Items)	8.0%	8.9%	5.5%	7.2%	4.7%	6.6%
Favorable Items						
2 QUIET	16%	9%	18%	10%	16%	9%
3 POPULAR	18	10	11	8	9	9
5 FULL OF FUN	12	19	9	8	9	9
8 NOT QUARRELSOME	2	10	7	7	15	7
10 TAKES CHANCES	14	24	4	6	2	5
12 NOT BOSSY	12	7	4	4	10	9
14 GOOD SPORT	4	7	5	12	11	7
16 NOT BASHFUL	8	10	10	9	9	8
17 GOOD AT GAMES	10	11	4	13	6	13
20 DOESN'T GET MAD EASILY	22	13	11	7	11	7
22 REAL BOY		23		12		14
24 ACTS LIKE LITTLE LADY	8		12		13	
26 GOOD-LOOKING			16	10	8	3
29 NOT A SHOW-OFF			10	5	9	6
30 LEADER			10	12	8	7
32 TIDY					13	7
35 DOESN'T FIGHT					2	5
36 FRIENDLY					9	7
38 TAKES JOKE ON SELF					6	4
Mean (Favorable Items)	11.5%	13.0%	9.4%	8.8%	9.2%	7.6%
Mean (All Items)	9.7%	11.0%	7.4%	8.0%	7.0%	7.1%

the group of negative items on which children voted more than on their favorable antonyms. It is of course possible that the children voted relatively less than the teachers on such items as FIGHTS and QUARRELSOME because of their general reluctance to make unfavorable characterizations of each other. But since these items, at least for boys, have positive correlations with REAL BOY, GOOD AT GAMES, and LEADER, this explanation seems inadequate. More probably the discrepancy between the relative emphasis by teachers and by children of certain items reflects a difference in their points of view as to which aspects of child personality are seriously undesirable.

The percentage of children named by the teacher shows a consistent decline over the range from Grade I to Grade V. This is true both of the average complimentary item (approximately from 12 per cent to 8 per cent) and of the average uncomplimentary item (approximately from 8 per cent to 5 per cent), but probably signifies no more than that the higher grade classes were larger. On individual items, grade-level differences are in most cases small. This finding is to be expected, for although adult observers have long noted behavioral differences between older and younger children, it must be remembered that these teachers were forced by the nature of the test to rate children on a relative rather than an absolute basis, and had, moreover, probably derived the subjective standards in terms of which they made their judgments from children falling within a narrow age range. Selecting items on which there is at all three levels a consistent difference in favor of one sex (irrespective of the statistical significance of the difference), one discovers that boys are more often mentioned by teachers on the items WIGGLY, NOT MANY FRIENDS, POOR SPORT, GETS MAD EASILY, TAKES CHANCES, and GOOD AT GAMES, while girls are more often mentioned on the items SERIOUS, DOESN'T TAKE CHANCES, QUIET, and DOESN'T GET MAD. The parallel between these results and those based on children's judgments is close, though there is less of a tendency among teachers than among children to assign all the favorable qualities to girls and their negative antonyms to boys.

D. SUMMARY OF CHAPTER IV

In this chapter the data have been analyzed in order to reveal characteristic patterns in (*a*) children's judgments of other children; (*b*) children's judgments of themselves; and (*c*) teachers' judgments of children. Marked differences in the volume of voting on different items were discovered in all three types of judgment data studied. With some exceptions both teachers and children made more nominations on favorable than on unfavorable items.

Grade-level comparisons of the volume of voting over the *test as a whole* did not show consistent trends, either with respect to self-mentions or mentions of others. However, the children became with advancing grade level more and more disinclined to name others as possessing seriously disapproved traits, whereas they tended to alter their judgments of themselves in the opposite direction (although even in the fifth grade favorable self-nominations were much more frequent than unfavorable ones). Teachers' judgments were not significantly influenced by the age of the subjects judged.

Sex differences in judgment patterns were noticeable. Girls voted more for others, but boys voted more for themselves. Also, girls cast more favorable votes for other girls than boys cast for other boys, but girls gave less favorable self-evaluations than did the boys. The latter difference between girls and boys is similar to that between older and younger children, and suggests that girls are more mature in their habits of judgment than are boys of the same grade level.

Children tend to cast considerably more than half of their votes for members of the same sex. However, certain qualities tend to be identified with boys and others with girls by all groups of informants. Boys are customarily identified on such items as WIGGLY, QUARRELSOME, FIGHTS, and SHOW-OFF, while girls are more often mentioned as QUIET, NOT QUARRELSOME, and GOOD-LOOKING. When children mention themselves, they tend to do so on items which are correlated with popularity for that sex—the girls choosing items connoting docility and amiability, the boys those connoting daring and sportsmanship.

CHAPTER V

AGREEMENT AMONG JUDGES ON THE REPUTATION TEST

THE Reputation Test is essentially a device for obtaining personality ratings of a child concurrently from many judges—from his classmates, from his teacher, and from the child himself. The judgments of these various informants reflect different frames of reference, and the determination of the degree to which the judgments are mutually congruent is of considerable theoretical and practical interest. In this chapter are reported findings on the agreement with which classmate judges characterize each other, and on the correspondence between pooled opinions of classmates and judgments from the special points of view of the teacher and the child himself.

A. AGREEMENT AMONG CLASSMATES

Perhaps the simplest index for evaluating consistency in judgment among classmates is the correlation between reputation scores on a given trait based on pooled judgments from a random half of the class and scores from the other half. A correlation of this type has been referred to by some investigators as the "reliability coefficient" of the trait concerned, a usage which has been avoided here because the term has also been applied to such distinct aspects of the test as the degree of overlap among the component items, and the temporal constancy of scores.

In point of fact, it has been shown that reputation testing compares very favorably with other types of rating procedures in "reliability" by either of the latter definitions.

Hartshorne, May, and Maller (3) reported for a group of fifth and sixth graders a correlation of .88 between half the items of a test of "altruism" and the other half. Their procedure was inappropriate in the present instance since our test was designed to tap as many different dimensions of reputation as possible.

Symonds and Jackson (8), whose test was designed to measure "adjustment" secured a retest correlation over a four-month interval of .64 using ninth graders as subjects. Tryon (10), using a test comparable to ours for a population of seventh graders, reports retest coefficients over a ten-day interval for different traits. These ranged between .45 and .95, only six out of forty being less than .65 and half of them being higher than .80. In the present investigation, systematic retestings were not feasible. However, correlations obtained for one fifth-grade class between individual test and group test scores on the various dimensions of the test averaged .77 over a two-month interval. (See above, Chapter II, B.)

In evaluating these findings, it must be recalled that retest coefficients on the Reputation Test are likely to be attenuated not only by temporal shifts in judgment such as affect all rating devices, but also by the sensitivity of the test to changes in the composition of the class group. For example, if at the first administration the class as a whole designates a given boy as most outstanding in sports and in the interval between tests a new child enters the class who is recognized to be an even better athlete, the score of the former will drop markedly since he will now receive no mention from those judges who make but one nomination each. Hence, the repeated demonstration of retest correlations as high as those cited above indicates that reputational status is surprisingly stable, at least over intervals of a few weeks or months.

The procedure used in the present study to measure agreement in class opinion was to divide each class alphabetically into two groups of judges composed of boys and girls in the same proportion. Correlations were then computed for each item-pair between scores received from one half of the judges and those received from the other half. In this analysis, the several classrooms at each grade level were pooled (necessitating correction for size of class), but scores received by boys and by girls were treated separately. The resulting coefficients, corrected by the Spearman-Brown formula, are presented in Table 10.

TABLE 10

Correlations Between Scores Received from One Half of Class and Scores Received from Other Half of Class
(Augmented by Spearman-Brown correction)

Item No.	Item-Pair	Grade I		Grade III		Grade V	
		Girls ($N = 49$)	Boys ($N = 70$)	Girls ($N = 340$, items 1–25) ($N = 178$, items 26–31)	Boys ($N = 352$, items 1–25) ($N = 194$, items 26–31)	Girls ($N = 306$)	Boys ($N = 322$)
1– 2	Wiggly–Quiet	.59	.61	.89	.88	.86	.88
4– 3	Not Many Friends–Popular	.76	.77	.86	.86	.92	.87
6– 5	Serious–Full of Fun	.46	.49	.72	.64	.71	.70
7– 8	Quarrelsome–Not Quarrelsome	.57	.40	.72	.82	.82	.83
9–10	Doesn't Take Chances–Takes Chances	.36	.64	.60	.75	.73	.80
11–12	Bossy–Not Bossy	.39	.51	.65	.65	.79	.85
13–14	Poor Sport–Good Sport	.65	.58	.65	.72	.76	.82
15–16	Bashful–Not Bashful	.51	.48	.60	.52	.84	.80
18–17	Not Good at Games–Good at Games	.70	.60	.76	.84	.83	.89
19–20	Gets Mad Easily–Doesn't Get Mad	.01	.63	.73	.67	.79	.83
21–22	Sissy–Real Boy		.63		.80		.88
23–24	Tomboy–Acts Like Little Lady	.76		.82		.89	
25	Best Friend	.70	.15	.57	.56	.40	.53
27–26	Not Good-Looking–Good-Looking			.77	.73	.89	.78
28–29	Show-Off–Not a Show-Off			.67	.74	.78	.87
31–30	Not a Leader–Leader			.31	.76	.75	.80
33–32	Untidy–Tidy					.83	.83
34–35	Fights-Doesn't Fight					.69	.88
37–36	Not Friendly–Friendly					.78	.74
39–38	Can't Take a Joke–Takes Joke on Self					.62	.62
	Mean r ("Best Friend" omitted)	.52	.58	.70	.74	.79	.82
	Max. S.E. (i.e., S.E. of zero r) for corresponding N's	.14	.12	.05 .07	.05 .07	.06	.06

As might be expected, the average correlation was much the lowest for the first-grade data ($r = .55$), but increased with the grade level of the subjects to .72 in the third grade and .81 in the fifth. While the differences were not statistically significant, there was a tendency for the reputation scores of boys to yield higher split-half-of-judges coefficients than those of girls. This should probably be attributed to the fact that boys received more negative mentions and hence showed greater spread in scores. The item-pairs Popular, Little Lady, Real Boy, and Quiet tended in most sets to show the greatest concurrence in classmates' opinion, perhaps because score variability was greatest on these variables. Best Friend was the only item (no pair) for which the correlations declined with increasing grade. This was to be expected, since the condition for a high r, that all judges name the same individuals, would demand in this case that children name their most popular acquaintances instead of their personal friends. That a few children nevertheless tended to garner a considerable percentage of all votes cast on this item was indicated by correlations as high as .40 and .53 for girls and boys, respectively, even in the fifth grade.

The correlations obtained are relatively high for a rating device and indicate a substantial degree of agreement among our judges in their identifications on the various items. Apparently the procedure of using many judges and requesting them to nominate only those whom they regard as outstanding compensates in considerable measure for the youth and psychological naïveté of the informants.

There remains the problem of determining the degree of actual disagreement among judges which is obscured by the practice of calculating a child's reputation score on a given dimension from the algebraic sum of the nominations he receives on the positive and negative items of the pair. Disagreement in this sense is defined as the mention of the same child for the possession of *both* the constituent traits of a dimension. Thus, partial disagreement is expressed concerning a child who receives five votes for QUIET and two votes for WIGGLY, but no disagreement is expressed concerning a child from the same class who receives three votes for QUIET and none for WIGGLY, although the algebraic sum of $+3$ is the same for both children. Complete disagreement occurs with reference to children who are mentioned equally often for the opposed traits of a dimension and who as a result obtain an algebraic score of zero.

Data on the frequency of complete disagreement on the average item-pair are presented in Table 11. Fluctuations from dimension to dimension are not marked, but the item-pairs on which the largest percentage of the group remained unidentified are naturally those on which score variability was least. Approximately one-fifth of the children received a score of zero on the average item-pair, the percentages rising from slightly more than 20 per cent in Grade I to almost 30 per cent in Grade III, and than falling to less than 20 per cent in Grade V. The proportion of those with zero scores who were simply not mentioned rises from around 45 per cent in the first grade to a high point of 70 per cent in the third grade. In the fifth grade, this proportion dropped to 50 per cent, though it should have been higher still, provided agreement among judges improves with age. However, the increased volume of voting in Grade V resulted in the mention of less deviate children, and concerning the proper identification of these, there would naturally be more disagreement among classmate judges. This interpretation of the trend is indirectly supported by the fact that cancellations contributed more heavily to the number having zero scores on SISSY–REAL BOY than on any other dimension, and on this item-pair voting tended to be especially heavy.

On the basis of our findings, one may state that on the average dimension, roughly one child in ten is completely disagreed about and a second child is ignored. However, these results are not as disturbing as they might seem, since complete disagreement is more likely to occur when a child is mentioned only once or twice on the opposed traits. Complete disagreement concerning a conspicuous child is very rare, and cases of children receiving as many as four votes on each pole hardly ever occur.

An approach to the measurement of concurrence in children's scores which indicates the amount of partial as well as of complete disagreement may be made simply by determining for each item the percentage of those mentioned for whom no dissenting votes were cast (i.e., for whom there were no mentions on the other item of the pair). This method is easily applied, but has the disadvantage of failing to take account of differences in the magnitude of individual scores. Children receiving 1 vote and 14 votes as WIGGLY are given equal weight, provided neither receives any votes as QUIET. (Children not mentioned on the item in question would of course not be considered to have been identified at all, since by the nature of the test a child receiving no votes can scarcely be said to have been "judged" even to be neutral.)

The standard of no dissenting votes as a criterion of agreement is a very stringent one,

TABLE 11

THE INFLUENCE OF CANCELLATION OF FAVORABLE AND UNFAVORABLE VOTES UPON THE FREQUENCY OF ZERO SCORES

Category	Grade I		Grade III		Grade V	
	Girls	Boys	Girls	Boys	Girls	Boys
1. Mean percentage receiving an algebraic score of zero (Mean for all item-pairs combined)	21%	22%	29%	28%	15%	17%
2. Mean percentage receiving no mention	12%	8%	21%	20%	8%	10%
3. Mean percentage receiving no mention of those with an algebraic score of zero (i.e., line 1./line 2.)	55%	40%	73%	68%	50%	56%

and Macfarlane, Honzik, and Davis (6) have proposed the modification of this criterion to include in the group "agreed about" those children who were positively identified on an item and who received not more than one dissenting vote. Naturally, children receiving exactly one vote on both the traits of the dimension would not be considered to have met even this criterion, since by the above definition they were completely disagreed about. Actually, the two criteria: (*a*) allowing no dissenting votes; and (*b*) allowing one dissenting vote, yield very similar results. However, the inclusion of cases receiving one dissenting vote increases by about 15 per cent the size of the group concerning whose proper identification there was concurrence among judges.

The percentages of boys and girls "agreed about" on the average favorable and the average unfavorable item are presented in Table 12.

In general, between half and two-thirds of the children are agreed about without serious dissent on the average item. Furthermore, agreement is somewhat better on positive than on negative items, for all groups of subjects save first-grade boys. This finding is to be attributed to the fact that more favorable than unfavorable votes were cast for all groups except first-grade boys. Since a child's chance of receiving a favorable mention was greater than his chance of receiving an unfavorable mention, he was less apt to receive a dissenting unfavorable vote when his score was favorable, than he was to receive a dissenting favorable vote when his score was unfavorable. A second finding, that boys (who received many more negative mentions) are agreed about relatively better than girls on unfavorable items and less well on favorable items, may be accounted for on the same basis.

With respect to age trends, agreement improves rather markedly from the first grade to the third, but shows no consistent change from the third grade to the fifth. One might have expected that the degree of concurrence would increase directly with the age of the judges, either on the assumption that older children are better observers or that they are more responsive to class-wide stereotypes of attitude. However, the lack of consistent improvement from the third to the fifth grade suggests that other factors, especially shifts in the general proportion of favorable to unfavorable voting, were sufficiently important to obscure any hypothetical relationship between the judges' accuracy and their age.

With respect to individual items, there are few which are consistently good or bad with respect to agreement for all groups of subjects. However, QUIET, POPULAR, and GOOD-LOOKING and their antonyms, elicit rather consistently good agreement, while on another small group including ACTS LIKE LITTLE LADY, REAL BOY, and GOOD SPORT and their antonyms, the agreement is for most groups of subjects relatively poor.

To summarize: the trends in score agreement parallel the trends in mean scores themselves. Girls, who tend to have more favorable mean scores than boys, are agreed about better than boys when identified for positive qualities, but are agreed about less well when identified for negative ones; and, as in the case of sex differences in mean scores, sex differences in agreement become less marked in higher grades. Approximately eight children in ten receive a score other than zero on the average dimension, and

TABLE 12

MEAN PERCENTAGE OF CHILDREN "AGREED ABOUT" ON FAVORABLE AND UNFAVORABLE ITEMS

Category	Grade I		Grade III		Grade V	
	Girls	Boys	Girls	Boys	Girls	Boys
Favorable items (No dissent)	46%	28%	57%	50%	55%	53%
Unfavorable items (No dissent)	27%	29%	43%	48%	39%	45%
Favorable items (One dissent)	62%	43%	70%	62%	70%	68%
Unfavorable items (One dissent)	39%	47%	56%	61%	53%	59%

from the third grade on, two-thirds of these are identified with substantial agreement. Hence, it seems safe to say that the Reputation Test yields sufficient data to make possible clinical characterization of somewhat over half the children on each dimension, and of a much larger proportion on the test as a whole.

B. AGREEMENT BETWEEN SELF-JUDGMENTS AND CLASS OPINION

The problem of studying agreement between self-judgments and classmates' judgments differs from that of analyzing concurrence within the pool of class opinion and requires other methods. In this section "agreement" is said to obtain when children not mentioning themselves on an item are not mentioned (or are mentioned only infrequently) by their classmates, and children mentioning themselves are mentioned by many of their classmates. Results have been analyzed for each item in terms of the critical ratio of the difference between (*a*) the mean reputation score from classmates of those who mentioned themselves and (*b*) the mean score of those who did not.

Biserial correlations were also calculated between self-mention vs. non–self-mention and class reputation for those items on which at least 5 per cent of the children named themselves. Unfortunately the infrequency of self-mention on some items not only precluded calculation of biserial coefficients, but also produced very large standard errors in the mean scores of the self-mention groups. For this reason a substantial absolute difference between self-mention and non–self-mention mean reputation score often failed to yield a significant critical ratio.

The critical ratio of the difference between mean scores of the self-mention and non–self-mention groups on the median item increased with age but was low in Grades I and III (.4 and 1.3, respectively), and reached a value of only 2.5 even in Grade V. Biserial correlations were also low, the median value for Grade V being close to .40. Sex differences were not marked, although critical ratios were slightly larger (indicating better agreement) for boys than for girls. For both sexes, there seemed to be a slight tendency toward better agreement on unfavorable items than on favorable ones. Although complimentary self-nominations were much more frequent than uncomplimentary ones, the latter were more likely to receive corroboration from the class.

With respect to individual items, there were none yielding consistently good agreement in all sex-age groups. Negative critical ratios (indicative of actual disagreement) were not uncommon in Grades I and III. Even in Grade V only about one-third of the items yielded positive critical ratios of 3.0 or more.

The general conclusion seems clear that children are either not very successful at observing in themselves characteristics which the class group identifies, or else are inhibited by self-protective mechanisms from expressing in this way their true opinions of themselves. In either case, self-nominations appear to contribute to the individual child's reputation record a type of data not duplicated by judgments from the classroom group.

C. AGREEMENT BETWEEN TEACHERS' JUDGMENTS AND CLASS OPINION

Since each teacher had completed the Reputation Test upon her own classroom, it was considered worth while to determine not only the quantitative aspects of teachers' voting patterns, but also the degree to which their judgments of individuals coincided with those of the children in their classes. To this end, critical ratios were computed to test the significance of the difference between the mean reputation score from classmates of those mentioned by the teacher and the mean score of those not mentioned.

Agreement between classmates' and teachers' judgments was considerably greater than that between classmates' judgments and self-evaluations, although in both cases agreement increased in direct relationship to the age of the group, and in both cases agreement was slightly better for boys than for girls (Grades III and V only). The tendency toward better agreement with the class on unfavorable items noted in the self-judgment data was not found in the teacher-judgment data.

With respect to individual items, agreement between teachers and classmates was highest on traits such as LEADER, BOSSY, and POPULAR, which the teachers were able to observe in class, and on traits such as WIGGLY for which the teachers probably determined the attitude of the class.

Agreement between teacher and class, expressed in terms of critical ratios, increased with grade level from a median value of approximately 1.2 in Grade I to 2.5 in Grade III and 3.4 in Grade V, but the corresponding biserial correlations reached a value of only .6 even in Grade V. Thus one may conclude that although by the fifth-grade level the teachers are identifying on most items a group whose mean score is significantly more extreme than that of the children not so identified, nevertheless there is too much discrepancy between the characterizations by teachers and by classmates to enable one to predict an individual child's status with the class from his reputation with the teacher.

In this connection, it should be emphasized that poor agreement does not necessarily indicate that either the teachers or the classmates are poor observers. The children and adults may have differed in their interpretations of the items themselves, and more important, were forming their judgments in different social contexts. For example, a significant *negative* relationship was found between teachers' and classmates' judgments as to which first-grade boys were BASHFUL, suggesting that shyness with adults may be a quite different matter from shyness with other children. It must be remembered that the teachers' contacts with the children were usually limited to the classroom, and that their opportunities to see the children in a wide variety of situations and thus to form judgments on many traits, were correspondingly curtailed. While the children, in neighborhood play, soon learned which were daring, good sports, etc., the teachers did not, and sometimes complained that on certain items they had had no opportunities to form judgments. For this reason, it is not surprising that teachers and classmates often disagree in characterizing individual children.

D. SUMMARY OF CHAPTER V

In this chapter have been summarized the findings from analyses of the agreement in characterization (1) among classmate judges, (2) between self-estimates and classmates' opinions, (3) between teachers' judgments and classmates' opinions.

Correlations between scores received from half the classmate judges and scores from the other half increased on the average item-pair from .55 in the first grade to .72 in the third and .81 in the fifth, indicating considerable agreement among classmates in identifying children on the various dimensions of the test.

On the average dimension, approximately eight children in ten received a definite characterization. Of the two children in ten receiving a score of zero, one was not mentioned at all and the other received equal numbers of favorable and unfavorable votes. Of those receiving a definite characterization on one item or other of the dimension, about 30 per cent in Grade I and 50 per cent in Grades III and V were identified without any dissenting votes on the opposed item (although the frequencies were increased by 10 per cent to 15 per cent if one dissenting vote was allowed). Thus it is apparent that the "dimension-score" (based upon algebraic cancellation of favorable and unfavorable votes) used in many of the analyses presented in this report was for about half the subjects something of an abstraction from the raw data on votes received. However, most disagreement occurred in the case of children occupying more or less median status in the group, and only rarely did children with markedly deviate scores receive dissenting votes.

Fluctuations in agreement in boys' and in girls' scores at the various age levels seemed to be in part a function of the total number of positive and negative votes received by the group of subjects concerned. Thus boys, who received relatively more votes on unfavorable items, were agreed about better than girls on these qualities, while girls were judged with more concurrence on the favorable items on which they received relatively more votes.

Agreement between reputation with a pool of classmates and *self-evaluations* (measured by critical ratios of the difference in mean reputation score of the self-mention and non–self-mention groups) was quite poor. Although agreement increased with grade level, the median critical ratio even in Grade V was only 2.5.

Agreement between classmates' judgments and *teacher nominations* was somewhat better than in the case of self-evaluations (C.R. of median item in Grade V = 3.4), but the teacher and the class often disagreed as to which individual children were most deviate on particular items of the test.

The results presented in this chapter indicate that judgments by classmates, teacher opinions, and self-nominations are by no means in perfect agreement with each other. However, the records of individual children discussed in Part II of this report indicate that such discrepancies in opinion are themselves meaningful. When interpreted qualitatively and configurationally, they make of the Reputation Test a richer and more sensitive means of diagnosing social adjustment than would a quantitative evaluation of classroom reputation considered alone.

CHAPTER VI

SUMMARY AND CONCLUSIONS

THIS INVESTIGATION of reputation differences was directed toward two major goals: first, the discovery of characteristic sex and age differences in the attitudes of elementary school children toward each other and toward themselves; second, the exploration of the possibilities of the Reputation Test as a clinical device for diagnosing the social adjustment of individual children. Part I of this report concerned the first of these goals. Findings on group differences presented in previous pages were analyzed not only for their significance as norms for diagnosis, but also for their reflection of the social demands to which the individual child must adjust.

A. DESCRIPTION OF TEST

The test used to explore these problems was constructed to measure children's attitudes toward each other on certain aspects of personality which they could readily observe in each other, which they could judge without emotional disturbance, and which were of significance to the clinician. In form, the test consisted of pairs of short questions, one item in each pair being considered favorable for adjustment and the other unfavorable. Examples are:

(+) "Which children are good sports and always play fair?"

(−) "Which children aren't very good sports, the poor losers?"

The children were asked to supply the names of those of their classmates who they thought best fitted each description. They were allowed to make as many nominations as they wished, to omit items if they could think of no one deserving mention, and to name themselves as often as they pleased. The teacher of each class also supplied data in this form on her pupils.

B. POPULATION STUDIED

The data upon which this investigation is based were obtained as a part of the Guidance Study of the University of California Institute of Child Welfare (5). The Reputation Test was administered in the public schools of Berkeley, California, in order to secure material on the social adjustment of these 250 Study cases, but since the test can be given only to an entire room, reputation data were secured also on the classmates of these children. Hence, the findings reported here are based on a much larger total population. The effect upon the sampling of selecting only classrooms containing Guidance Study cases is unknown; but the size of the group and the fact that the testing was carried on in all the elementary schools of Berkeley, make it seem likely that the population was quite typical of the entire elementary school population of that city during the years the data collection was in progress (1936-1939). Classrooms in all six grades were tested, but the group analyses reported here were performed only at the alternate grade levels I, III, and V. For Grade I, 119 children in four classrooms are included; for Grade III, 692 children in 20 classrooms; for Grade V, 628 children in 18 classrooms. Boys constituted 60 per cent of the Grade I sample, but in Grades III and V, boys and girls were approximately equal in numbers.

C. ADMINISTRATION AND SCORING

For the first four grades, the children's opinions were obtained in individual interviews. However, an experimental testing of a fifth-grade class indicated the feasibility of group administration at this level, and in Grades V and VI the test was given to an entire class at one time, and the children requested to write down their nominations in test booklets. The number of mentions given to and received from others, self-mentions, and mentions made by the teacher were tabulated by sex and grade and used in the various analyses.

To obtain the reputation scores for each child, the number of mentions on each item (excluding self-mentions) was divided by the number of classmates who might have mentioned him. Scores on each dimension (composed of the positive and negative items of each pair) were obtained by dividing the algebraic sum of positive and negative mentions by the number of classmates.

D. Group Differences in Reputation

So far as votes *received* are concerned, it was found that on the majority of items girls enjoy much more favorable status in the eyes of their classmates than do boys. The average girl was judged to be QUIET, POPULAR, FULL OF FUN, NOT QUARRELSOME, a GOOD SPORT, a LITTLE LADY, GOOD-LOOKING, NOT A SHOW-OFF, TIDY, and FRIENDLY. The average boy was mentioned on the items WIGGLY, QUARRELSOME, BOSSY, SHOW-OFF, TAKES CHANCES, NOT BASHFUL, GOOD AT GAMES, and REAL BOY. Although aggressiveness was regarded as typical masculine behavior, the unfavorable connotations of such qualities as BOSSY, SHOW-OFF, QUARRELSOME, etc. suggest that boys are subject to much more disapproval than are girls in learning their appropriate sex role.

Age differences in reputation were less marked than the sex differences cited above, but a tendency with increasing grade level for girls' scores to grow less favorable and for boys' scores to improve, operated to reduce the magnitude of the sex differences. The scores of both boys and girls were determined primarily by the opinions of judges of the same sex. Nevertheless, in the first grade, girls' scores were determined relatively more by boys than vice versa, while by the fifth grade, the situation was reversed and boys' scores were determined relatively more by girls. Since sex differences in reputation were so conspicuous even in the first grade, and since the picture for girls was so markedly favorable and for boys so markedly unfavorable, it seems likely that these findings reflect not only behavioral differences between boys and girls, but also the stereotype that little boys are aggressive and dominant, little girls docile and well-behaved.

E. Group Differences in Judgment

Marked differences in the volume of voting on different items were found in teachers' mentions and in children's mentions of each other and of themselves. Both teachers and children made more nominations on favorable than on unfavorable items, and insofar as children's judgments of each other are concerned, this tendency grew more pronounced in higher grades. However, the trend for self-mentions was in the opposite direction, self-judgments being made almost exclusively on favorable items in the first grade, but only slightly more frequently on favorable than on unfavorable items in the fifth grade.

Older children voted more heavily for others (and markedly more for themselves) than did younger children, but at all grade levels considerably more than half of the judgments concerned persons of the same sex as the judge. Certain qualities, especially WIGGLY, QUARRELSOME, FIGHTS, and SHOW-OFF, were customarily attributed to boys by judges of both sexes, while others such as QUIET, NOT QUARRELSOME, and GOOD-LOOKING were more frequently used to describe girls. Self-mentions were made most often on items correlated with popularity for children of the sex concerned, the girls choosing items connoting amiability and docility, the boys those connoting athletic skills and daring.

In several respects, the judgment habits of girls and of boys stand in contrast: girls voted more for others than did boys at all three levels studied, but boys made more self-evaluations than did girls. Also, girls judged others of the same sex more favorably than did boys, but gave less favorable self-evaluations. Since these differences between boys' and girls' habits of judgment parallel those between younger and older subjects, the tentative conclusion is offered that girls are more "mature" in the pattern of judgments they make than are boys.

F. GROUP DIFFERENCES IN AGREEMENT AMONG JUDGES

In order to measure the pervasiveness of opinion among classmate judges, the latter were divided into two halves and correlations computed between item-pair scores received from one half, and scores received from the other half. These "agreement" coefficients increased on the average item-pair from .55 in Grade I to .72 in Grade III and .81 in Grade V, indicating considerable community of group opinion.

Disagreement in classmates' judgments was extremely infrequent concerning children whose scores differ markedly from the average, but was very common in the scores of children who received few votes. On the average dimension about 20 per cent of the children received a dimension score of zero, and of these about half were completely disagreed about (i.e., received equal numbers of favorable and unfavorable votes). Of those receiving a definite characterization on one item or the other of the average dimension, approximately 30 per cent in Grade I and 50 per cent in Grades III and V were identified without any dissenting mentions on the opposite item. If those receiving only one dissenting vote are included in the group agreed about, the proportions are increased by about 15 per cent. In general, there was more concurrence in the scores of boys identified on negative items, and in the scores of girls identified on positive items.

Agreement was quite poor between self-mentions and scores received from a pool of classmates, although the agreement improved with the age of the subjects concerned. Agreement between classmates' opinions and the teacher's judgment was somewhat better, but even here it was not until the fifth grade that the mean score of children mentioned by the teacher on the average item was significantly greater than the mean score of children not so mentioned. Hence, self-nominations, teacher mentions, and classmate judgments cannot be considered equivalent in the information they yield concerning the individual child. However, all may be useful in evaluating the adequacy of his social adjustment and the social climate to which he must adjust.

II. THE DIAGNOSIS OF SOCIAL ADJUSTMENT

CHAPTER I

THE DIAGNOSTIC USE OF THE REPUTATION TEST

IN THE preceding paper, the data from an extensive administration of the Reputation Test were analyzed for the purpose of revealing sex- and age-group differences in children's attitudes toward one another and toward themselves. A second major purpose of the investigation was to explore the usefulness of the Reputation Test as an instrument for studying the social adjustments of particular children. From this point of view, results of the group analyses may be regarded as norms for clinical apppraisal. The successful application of group findings to the individual is a *sine qua non* for any device claiming clinical utility, despite the unavoidable loss in statistical precision when the focus is shifted from the central tendency to the single case. The present paper will therefore be devoted to exploring the clinical potentialities of the Reputation Test as a device for diagnosing individual children's social adjustment to their peers.

A. THE CLINICAL SIGNIFICANCE OF A CHILD'S REPUTATION

We have assumed that a child's reputation *with* his peers is of central importance in assessing his social adjustment *to* those peers. Hence, classmates' judgments constitute the most important basis for diagnostic interpretations of Reputation Test results. The evaluation of these judgments, however, can be made on different levels of refinement.

At the most general level, one can simply inspect the over-all record to note the number of mentions received by the child in question, and the ratio of favorable to unfavorable nominations. This method is quick and is probably adequate to locate both those who are so actively disliked, and those who are so utterly ignored as to constitute the most serious problems of adjustment. In Grades I and II, the reliability of scores on individual traits is so low as to suggest that more refined methods are probably unjustified.

A somewhat more detailed approach involves using the profile of a child's scores on specific dimensions of the Reputation Test to discover the particular favorable and unfavorable traits which his classmates perceive in him. This type of analysis, by indicating specific areas of maladjustment, often helps give direction to remedial efforts. In the present study, score reliabilities were high enough in our judgment to justify such analysis for the third and higher grades. The primary limitation of this approach is the restriction imposed upon the manifest content of the test by the fact that the judges are children. It is necessary not only to keep the test relatively short and the vocabulary simple but also to avoid items, despite their psychological importance, which might make the children anxious or defensive. The test used in the present investigation proved satisfactory in these respects and indicates the range and number of items which can be included.

An interpretation much richer and more differentiating than that based on a score profile alone can be made by utilizing in addition the many qualitative cues afforded by a record as a basis for inferences which go beyond the manifest item content of the test—for example, taking into account discrepancies between the teacher's judgments and those of the class, differences in general reputation with boys and with girls, disagreement among judges on particular items, etc. This inferential level of interpretation demands clinical judgment and is subject to error and bias on the part of the interpreter. However, the variety of data available makes it possible to check one's inferences for their congruence with independent aspects of each record as well as for their consistency with each other. The case interpretations to be presented in the next chapter are illustrative of the potentialities and limitations of this approach.

In the present investigation, the features of the child's reputation record that were inspected routinely for diagnostic leads included the following:

1. The number of children who mentioned the subject one or more times in the course of the test. This afforded a rough index of the child's conspicuousness in the group—the popular, as well as the actively disliked, falling at one end of the scale, the isolated and withdrawn at the other. Modifying factors included the proportion of boys and of girls who named the subject and the attendance record of the child himself. Children's memories for absent classmates are very short, and inconspicuousness on the part of someone who is frequently absent may have no special significance.

2. The total favorable-unfavorable balance of nominations received by a child on all items of the test. This balance supplied a sort of reputational base line for the child concerned, but had to be considered in relation to the sex-grade norm, since girls ordinarily have much more favorable total reputations than boys.

3. The child's reputation with classmates on the various traits included in the test. Since not all variables were equally important in the children's eyes, evaluation of a child's score profile had to take account of the relationships between particular items and the children's conceptions of maturity, popularity, masculinity-femininity, etc. (See Studies in Reputation I, Chapter III, and [12].) Another factor modifying the significance of a particular score was the amount of disagreement involved, both between boy judges and girl judges, and within either sex group. Marked disagreement was relatively rare but usually significant.

4. The number of boys and of girls naming the subject on the BEST FRIEND item. Sociograms based on the voting on this item were constructed for each classroom, and used to evaluate the child's position in the social organization of the group (13). These sociograms provided a check on inferences as to the child's acceptance or rejection by his classmates, and frequently resolved questions as to the meaning of disagreement on other items by relating such disagreement to the clique membership of the judges concerned.

5. The record of the teacher's nominations of the child. The teacher's mentions were by no means perfectly correlated with those of the children, and constituted an independent set of judgments made from an adult point of view. Differences in emphasis between teacher and classmates often made possible inferences about the child's maturity, docility, and orientation toward the teacher's goals.

B. Projective Aspects of the Reputation Test

The aspects of the Reputation Test discussed thus far concern judgments made about a given child by others—his classmates and teacher. Additional diagnostic sources are the judgments made by the child himself. Utilization of this "projective" aspect of the record follows from the fact that a person's perception of others depends not only upon their actual behavior, but also upon the psychological needs and defenses of the person himself. In evaluating a child's reputation with others, the fact that each judgment made about him involves the perceptual bias of the judge is not a serious limitation because a large number of judges are pooled and idiosyncratic votes tend to cancel out. In evaluating the child's own voting record, "bad" self-nominations, i.e., those which contradict group opinion, can be a source of insights into the child's own areas of insecurity and his modes of defense. Theoretically, a child's "bad" judgments of others could be used in the same way, but since in the present study corroborative clinical material was available for only a few children in each room, voting idiosyncracies were evaluated only on the quantitative level, i.e., by noting items omitted in judging others and items voted on much more heavily than usual.

In order to differentiate between omissions which resulted from haste, distraction, or turning two pages at once, and those which were indicative of an emotional reaction to certain items, a clinical appraisal of the total record was required. If such omissions were related to the child's own reputation, we tended to regard them as psychologically significant. Most of the accidents of test administration operate to reduce the number of mentions made by a child. In instances where a child casts an excessive number of votes on a particular item, as for example, the little girl who named herself as a TOMBOY and all the others as LITTLE LADIES, it seems reasonable to conclude that that item has special importance for the child concerned. Even in such instances, consideration of the relative instability of a single child's voting record suggests that projective interpretations of a voting pattern should be made cautiously and only in relation to other aspects of the record which can provide independent checks.

The features of a child's voting which we attempted to relate to the personality picture afforded by classmates' judgments included the following:

1. The number of classmates mentioned by the subject one or more times on the test as a whole, as compared with the number mentioning him. This aspect of the record gave some hint as to the breadth of the child's acquantanceship. Especially important was the ratio of boys to girls mentioned. Children ordinarily named more of their own sex. A tendency to mention more of the opposite sex was quite rare and often indicative, especially among boys, of exclusion from the play group of the same sex.

2. General aspects of the voting record—the total volume of voting, the distribution of nominations between boys and girls, and the ratio of favorable to unfavorable votes. These aspects of the record often supplied cues as to the child's general atttiude toward other children of the same and of the opposite sex.

3. The distribution of a child's votes among the various items. The child's reputation with others on items which

he omitted or emphasized by heavy voting was noted for hints as to areas of defensiveness, etc.

4. The child's voting on the BEST FRIEND item. This item yielded leads as to the child's awareness of clique groupings, and of his own position with respect to the cliques. Secure clique leaders rarely mentioned more than three or four friends. Among those who received no votes on BEST FRIEND, at least two patterns emerged—some naming several class leaders, others responding to their own isolation by naming only one other child placed in a similar fringe position, or none at all.

5. The subject's mentions of himself on various items. The total number of self-nominations, the particular items on which they occurred, and the total balance of favorable to unfavorable self-nominations were inspected for the light they might throw upon the child's conception of himself. Areas of tension were sometimes suggested by defensive avoidance of self-nomination on unfavorable items for which the child's reputation with others was extreme, or, among younger children, by self-nomination for the opposite trait.

In summary, the child's voting pattern was compared with his own reputation, in a search for clues that would permit interpretation to go beyond the manifest content of the Reputation Test. It deserves emphasis that discrete elements of the record can gain meaning in relation to other such elements. Successful interpretation means the extraction of full and valid meaning from the particular combination of elements observed.

C. NORMS

There remains the problem of the appropriate choice of reference norms for clinical appraisal. Since the means and standard deviations of the various scores and indices differ markedly for different sex and age groups, it is apparent that the significance of a given number of votes received or given on a particular item cannot be gauged accurately by its deviation from zero.

Better alternatives are either norms for the grade and sex group based on several classes, or standards derived from the particular classroom in which the child happens to be. Obviously, the former are the more stable and accurate since they are based on many more subjects. In the present study, however, the various classes of a given grade level were located in different schools and are known to have differed as to the socio-economic background of the pupils and perhaps as to a variety of personal standards and attitudes correlated with this difference. It is possible, moreover, that there were slight differences in administrative procedure making for heavier voting in some rooms and lighter voting in others. Insofar as these factors enter into the results, the norms for a given room constitute a more adequate basis for describing individual deviations than do those based on a composite of findings from all the rooms of a given grade level. For these reasons an attempt has been made to relate the individual's record *both* to the grade norms and to his own classroom.

On the individual record tables presented in the next chapter, such of a child's mentions or scores as are most extreme among children of the same sex in his particular room have been printed in **boldface** type, those which are next to the most extreme in *italics*. Each record was also evaluated by comparison with the appropriate grade-sex norms derived by pooling classrooms, but to save space, such norm values have been omitted from the record tables except as they are implied in the dimension standard scores. The significance of particular self-nominations and teacher nominations may be judged by reference to the percentage frequency values in Studies in Reputation I, Tables 8 and 9, respectively; the significance of votes given to classmates by reference to the grade-sex means of Table 6.

CHAPTER II

ILLUSTRATIVE CASE RECORDS

A. INTRODUCTION

THE PRESENT chapter is devoted to diagnosing the social adjustments of five children on the basis of their Reputation Test records alone, and comparing the results with relevant case material from other sources. All subjects are members of the Guidance Study, and took the Reputation Test either in the third grade (Cases I and II) or in the fifth (Cases III, IV, and V).[1] They have been selected from among those available in order to illustrate different types of Reputation Test records and the clinical findings which may be associated with each. Case I, Mary, is a third grader whose Reputation Test interpretation of excellent social adjustment is in agreement with the extensive case history material compiled by the Guidance Study. In sharp contrast stands Alice, Case II, with poor adjustment indicated by the Reputation Test and confirmed by the case data. Case III, Margaret, has been included to show some of the limitations of our approach, for although the Reputation Test indicates correctly her poor adjustment, it is inadequate for specifying the nature of her problems. Case IV, Fred, shows a pattern of moderately good adjustment in a fifth-grade boy, which is confirmed by the case history. His test profile correctly forecasts the nature of his later dislocation. Case V, James, is included because his record illustrates the capacity of the Reputation Test to disclose difficulties in peer-relationships which may escape for a time the notice of adults.

It should be made explicit that these case interpretations constitute illustrations of the diagnostic potentialities of the Reputation Test rather than a "validation" of it. Not only are the case records few, but they are more deviate and supply more clues for interpretation than would a similar series collected at random from among the entire group tested. Indeed it would not be possible to write a full or differentiating description of children whose patterns lie near the group norm.[2] Further, it is not possible to be sure that the case material included here gives an unbiased picture of the children in question, since the total volume of life history data accumulated from birth is very large, and only condensed excerpts from it could be included here. To minimize the degree of distortion from this source, the case summaries were written without reference to the Reputation Test interpretations and prepared by a routine selection from the same sources for each child.

For each case, material will be presented in the following order:

1. The Reputation Test record. Owing to the expense of reproduction, the classroom sociogram and sex-age group norms have been omitted. Those of the child's

---

[1] First graders have been omitted because of the unreliability of their scores. Second, fourth, and sixth graders have been omitted because many of the group analyses were not undertaken at these levels. However, the consistency of age trends makes it seem likely that norms for these grades obtained by interpolation from adjacent levels would be accurate enough for practical clinical appraisals.

[2] This is not a serious limitation upon the clinical usefulness of the Reputation Test to the teacher or school counselor, since children whose social adjustment is poor do not ordinarily have average records, being either recognizably disliked, or isolated and unnoticed to a degree which is itself suspicious.

scores which are most deviate in his own classroom group are printed in **boldface**, those which are second most deviate in *italics*.

2. An "interpretation" of the record, i.e., a description of his probable personality characteristics and defenses in the area of social adjustment, based both upon his reputation with others and upon his own voting.

3. Clinical data extracted from the case material collected nearest the time when the Reputation Test was administered.

4. The writer's comment upon the degree of agreement between the Reputation Test findings and the case material.

Case numbers and names are fictitious, and facts which might be identifying have been disguised.

B. Case I: Mary, a Third-Grade Girl

1. *Reputation Test Interpretation*

Mary is a member of a large third-grade class composed of approximately equal numbers of boys and girls and located in a school drawing from a rather prosperous middle-class neighborhood. She completely dominates the social organization of the girls, eleven of them and even one boy naming her as Best Friend. Even more unusual is the fact that she in turn mentions as her friend not only a girl but also a boy (the most popular one), though nominations of boys by girls are extremely rare at this age.

Mary is named by, and in turn names, many more other children on remaining items of the test than is usual for third-grade girls, but this result is hardly surprising in view of her central position in the group.

As might be anticipated considering her wide popularity, her reputation on other qualities is so extremely favorable that one suspects a marked halo in some of her scores. On all items save Quiet and Not Bossy, she receives from girls more favorable votes than any other girl in the class. The fact that she receives from boy judges more mentions than any other girl on 6 of the 12 favorable items, and the second most on 2 more, is quite as unusual. Some idea of the relative infrequency of such a record as compared with the scores of all the third-grade girls may be gained from her standard scores on a few item-pairs on which she is especially deviate: Best Friend (5.0σ), Good at Games (4.4σ), Popular (3.9σ), Full of Fun (3.4σ), and Good Sport (3.3σ).

However, on most dimensions there are one or two children who mention her for the unfavorable quality in question, indicating that at least a few have not fallen under her spell. Some suggestion of the reason for this lack of unanimity is afforded first, by the fact that she receives the largest score from girls for Bossy (seven such votes, three dissenting votes); second, by the fact that boy *and* girl judges mention her more frequently than any other girl *both* on the Gets Mad Easily and on the Doesn't Get Mad items, a very unusual degree of disagreement.

Her teachers' mentions of her as Popular, Not Bashful, and Little Lady are not out of the ordinary.

The pattern of Mary's mentions of others is lacking in significant detail, save that the only item-pair which she omits entirely is the Gets Mad–Doesn't Get Mad dimension, on both poles of which she receives extreme scores from her classmates. This circumstance suggests that she may feel a little uncomfortable over this one flaw in her own near perfect social technique and prefer simply to skip the subject altogether.

Mary's four self-mentions (Popular, Full of Fun, Not Quarrelsome, and Good at Games) are all favorable. This result is not unusual among children of this age, and all four are items frequently so used by third-grade girls. However, it is interesting that on the four on which she names herself she enjoys a markedly favorable score, whereas, she fails to mention herself at either pole of dimensions on which her status is less assured. It seems possible that this child, whose social skill and awareness of others is obviously so marked, has made an honest statement of her assets, but prefers to appear in as good a light as possible and simply omits self-judgment with reference to traits on which her behavior is less exemplary.

Summary based on reputation data. Mary enjoys the approval of her classmates to a more marked degree than anyone else included in this investigation. The resulting "halo" in judgments concerning her makes it difficult to characterize her with reference to specific qualities. However, she is apparently buoyant, a good sport, a little lady, and recognized as good at games even by boys. On the other hand, her dominant position is apparently resented by some children, and by them she is regarded as bossy and prone to get mad. She mentions her principal assets quite accurately, but avoids mention of any unfavorable trait. The fact that she omits the Gets Mad–Doesn't Get Mad dimension altogether

TABLE 1

REPUTATION TEST RECORD OF CASE I, MARY

Sex—Female	Age—8 yrs., 4 mos.	Grade—Low third
Girls Present in Class —20.	Proportion Mentioning Mary—95%	(Norm = 47%).
Boys Present in Class —19.	Proportion Mentioning Mary—84%	(Norm = 34%).
Girls Enrolled in Class—21.	Proportion Mary Mentioned —65%	(Norm = 54%).
Boys Enrolled in Class—21.	Proportion Mary Mentioned —52%	(Norm = 40%).

Item-Pairs Negative-Positive	Votes Received						Votes Given					Item Symbols
	From Girls		From Boys		Standard Score	From Teacher	To Self	To Girls		To Boys		
	Neg.	Pos.	Neg.	Pos.				Neg.	Pos.	Neg.	Pos.	
WIGGLY-QUIET	2	*4*	1	3	.4			1	1	1	1	Q
NOT MANY FRIENDS-POPULAR		16		7	3.9	+[a]	+	1	3	1	3	P
SERIOUS-FULL OF FUN		9		5	3.4		+	1	2	1	2	FF
QUARRELSOME-NOT QUARRELSOME	1	8	1	5	2.6		+	2	2			Ql −
DOESN'T TAKE CHANCES-TAKES CHANCES	1	9			3.0			1				C
BOSSY-NOT BOSSY	**7**	3	2	1	−1.5			1	1	1		B −
POOR SPORT-GOOD SPORT	1	11		*3*	3.3				4	2	1	S
BASHFUL-NOT BASHFUL	2	7	2	3	1.6	+			1			Ba −
NOT GOOD AT GAMES-GOOD AT GAMES		12	1	7	4.4		+	2	1		2	G
GETS MAD EASILY-DOESN'T GET MAD	**4**	6	3	*3*	.3							M −
SISSY-REAL BOY	..	..	..	..	..	..	..	..	..	2	2	RB
TOMBOY-LITTLE LADY	1	12	2	9	2.6	+		1	5	..	..	LL
BEST FRIEND (no pair)	..	11	..	1	5.0			..	1	..	1	BF
Totals	19	108	12	47	..	0 − 3 +	0 − 4 +	10	21	8	12	

Note—Boldface entries indicate child was most deviate among classmates of same sex; italic entries indicate child was next to most deviate.

[a] A plus entry indicates a mention on the positive item of the pair; a minus, a mention on the negative item.

even in judging others suggests, moreover, that she may be a little uncomfortable over her failure to measure up to her customary level on this dimension. However, she is an unusually successful child, and very well adjusted in the domain of social interrelationships.

2. *Excerpts From the Case History*

a. Physical development. Mary is average in height, weight, and general development.

b. Mental test results. Stanford-Binet IQ = 148. (Other results have varied from 137 to 157.) Examiner's comments: "Keen, alert, definite drive for success, but not a show-off. Poised, modest, has good sense of humor. A charming personality quite in harmony with its environment."

c. Comments of teacher. (Data obtained five months before Reputation Test.) "Extremely friendly, likes everyone. Popular both with girls and boys. Very stable and grown up, gay and enthusiastic. Good-natured. Would never pick a quarrel—acts as arbitrator for differences between other children. More independent than average—plays with others but not dependent on group. Enjoys excelling but not upset by failure. Does excellent school work. I have never had such a well-liked child before."

d. From an interview with Mary at eight years. (With reference to social adjustment): "I like all the girls. The boys chase me but I don't tell on them. I don't like to quarrel—when I do, we're friendly again in about five minutes, so what's the use?"

e. Comments of case worker. (Based on interviews with child and with parent.) "Has four close friends and many more with whom she is less intimate. Is well liked by children and adults. Knows she is well liked but does not seem quite secure in this direction. Would like to be better in everything. Has tendency to sensitivity. Unusually self-confident and well poised but decidedly restrained and self-composed. Very capable; sweet and passive manner but excellent at managing others. Is neat and almost piquant in appearance, but not good looking. No real problems noted."

3. *Comment*

In this case the Reputation Test results are in good agreement with the data furnished by the case history. Moreover, the Reputation Test indicates the presence of some less desirable qualities in Mary's personality (BOSSY and GETS MAD), despite the fact that one might expect that child judges would be influenced by a generally favorable set to report too bland a picture of her. That the unfavorable judgments passed on her and the suggestion of inner tension may have a basis in fact is indicated by the case worker's comment that she "has a tendency to sensitivity," and again "knows she is well liked but does not seem quite secure in this direction." Obviously this is a very unusual child and a correspondingly unusual record. The psychologically more significant problem of how she acquired such a personality is unfortunately beyond the realm of the Reputation Test.

C. Case II: Alice, a Third-Grade Girl

1. *Reputation Test Interpretation*

Alice is a member of a classroom containing 14 girls and 18 boys. In this group she receives no mentions as a Best Friend. She herself names but one girl, who in turn votes only for the class "star." The total number of children mentioned by Alice on the entire test is not unusual, although she mentions relatively more boys and fewer girls than the average third-grade girl. The number of girls mentioning her is near the norm, but she is mentioned by about twice as large a proportion of boys as usual.

Alice is mentioned more often by boy judges than any other girl in the group on all but three of the unfavorable items included in the third-grade test, and on two of the remaining ones (Quarrelsome and Bossy) she receives the second most deviate score of any girl. Girls in Alice's class mention her most frequently on three of the traits for which boys characterize her as extreme (Not Many Friends, Serious, and Doesn't Take Chances), and while the girls mention her on other negative items, her poor reputation is less extreme with them than with the boys. As compared with third-grade girls in general, Alice's most extreme standard scores are on the negative poles of item-pairs, e.g., Doesn't Take Chances (−4.9σ), Not Good at Games (−3.8σ), Serious (−2.7σ), Bashful (−2.7σ), and Poor Sport (−2.4σ). Only on the item Little Lady does Alice obtain a positive standard score (1.3σ), but 7 of the 8 votes she received were cast by boys. In the third grade this item is correlated more with "submissive" variables than with Popular, and by the fifth grade conveys an implication of timidity and incompetence at games. Hence, Alice's score on this so-called "positive" item is at best indicative of her possession of a rather unvalued quality, and corroborates the evidence of her other scores that she is serious, timid, lacking in either motor or social skills, and in consequence isolated and unpopular in her group.

Alice receives 6 votes from her teacher, viz.,

TABLE 2

Reputation Test Record of Case II, Alice

Sex—Female Age—7 yrs., 10 mos. Grade—Low third

Girls Present in Class —14. Proportion Mentioning Alice—38% (Norm = 47%).
Boys Present in Class —18. Proportion Mentioning Alice—72% (Norm = 34%).

Girls Enrolled in Class—14. Proportion Alice Mentioned —46% (Norm = 54%).
Boys Enrolled in Class—18. Proportion Alice Mentioned —50% (Norm = 40%).

Item-Pairs Negative-Positive	Votes Received						Votes Given					Item Symbols
	From Girls		From Boys		Standard Score	From Teacher	To Self	To Girls		To Boys		
	Neg.	Pos.	Neg.	Pos.				Neg.	Pos.	Neg.	Pos.	
Wiggly-Quiet			**2**	1	− .6				1	1		Q
Not Many Friends-Popular	**3**		**3**		−1.7	−		**2**	*4*			P
Serious-Full of Fun	**3**		**5**		−2.7						1	FF
Quarrelsome-Not Quarrelsome			*1*		− .7							Ql –
Doesn't Takes Chances-Takes Chances	**5**		**6**		−4.9	−			1		2	C
Bossy-Not Bossy	1		*3*	1	−1.1			1	1			B –
Poor Sport-Good Sport	2		**4**		−2.4	+	+		*4*	**2**		S
Bashful-Not Bashful	1		**6**		−2.7							Ba –
Not Good at Games-Good at Games	*3*		**9**		−3.8	−	−		2		1	G
Gets Mad Easily-Doesn't Get Mad			**3**		−1.2	+			1			M –
Sissy-Real Boy	..	..	..	..	..	..	..	..	..		2	RB
Tomboy-Little Lady		1		7	1.3	+	+		2	..	..	LL
Best Friend (no pair)	..		..		−1.1			..	1	..		BF
Totals	18	1	42	9	..	3 − 3 +	1 − 2 +	3	17	3	6	

Note—Boldface entries indicate child was most deviate among classmates of same sex; italic entries indicate child was next to most deviate.

Not Many Friends, Doesn't Take Chances, Not Good at Games, Good Sport, Doesn't Get Mad, and Little Lady. In view of Alice's almost unrelievedly unfavorable reputation with her group, it is unnecessary to say that the teacher's negative statements about Alice agree with class opinion, while two of her positive ones (Good Sport and Doesn't Get Mad) are opposite to the class judgment. It is possible that the class mentions her on the undesirable poles of these variables because of a general unfavorable attitude toward her, but since they have better opportunities to observe such behavior than does the teacher, their attitude cannot be dismissed. In any event, Alice's reputation in this group is an unenviable one, and is as nearly the antithesis of Mary's as could be imagined.

Alice mentions herself as a Good Sport and a Little Lady, but concedes that she is Not Good at Games. In voting for other children, Alice gives more favorable than unfavorable mentions both to boys and to girls, and somewhat fewer unfavorable mentions to either sex than is usual for third-grade girls. This result is unexpected, since one might anticipate that a child who is so markedly rejected by her group might in turn reject them and distribute numerous uncomplimentary mentions.

The only items on which she casts an unusually large number of votes are Not Many Friends and Popular for girls, Good Sport for girls, and Bad Sport for boys, suggesting that these topics may be important to her. One wonders if her voting for boys on the Bad Sport item, considered together with the large number of unfavorable mentions she received from boys, reflects the possibility that they tease her a good deal. Certainly the boy judges are more hostile than the girls. Over the test as a whole she makes nominations on only four of the unfavorable items, but uses all except two of the favorable ones. These facts taken together suggest that Alice is not actively rejecting her classmates or their values, but lacks the ability to win friends among them.

Summary based on reputation data. The total picture is that of a very much inhibited and shut-in personality, though she may react aggressively if teased or intruded upon. The fundamental reasons for her generally low status in the group are not obvious, and it seems unlikely that poor motor skills alone are enough to place a girl in such an undesirable position. Although there is probably much negative halo in her record on specific items, the reputation findings strongly suggest the advisability of some sort of psychological assistance. Probably even this has little chance to succeed while she remains in this classroom where opinion has already crystallized against her.

2. *Excerpts From the Case History*

a. Physical development. Alice is slightly above average in height, weight, and general development.

b. Mental test results. Stanford-Binet IQ = 106. (Has varied from 102 to 114.) Examiner's comments: "An attractive looking child, socially at ease. Too much surface for dependability. Bluffs, guesses when she isn't sure. Dislikes hard tasks and stalls."

c. Comments of teacher. "Used to be very unpopular but now liked better and tries to be friendly. Disappointed if not accepted in group. Never quarrels. Tense and excitable. Least self-reliant in class. Likes praise. Behavior is on the whole better than average. Making remarkable progress in attitude toward work and work itself. Mother waits on her and won't let her go to or come from school alone."

d. From an interview with Alice. "Kids are mean to me—they tease and won't play. I invite them to my parties—then they ask me, but when I get there they just take my present and say 'goodbye'—don't even ask me in."

e. Comments of case worker (based on interviews with child and parents). "Strong and unwholesome emotional dependence on mother, which is perpetuated by the latter. Insecure relationship with father who is moody and undemonstrative. Poor social techniques and poor social adjustment. Not accepted by group and disliked by most children. Alice is careless of property and quarrels much with her mother and playmates. Is teased by others because they can always get a rise. Mother dresses her and undresses her to prevent temper tantrums in which she smashes objects and screams, sometimes for hours. Defiant and sullen if crossed. Irritable if plans upset—adjusts with difficulty. Aggressive, won't follow others, is angry if others don't follow her. Very timid. Whines and is fretful. Very mercenary—refuses to help in any way unless paid. Child badly spoiled."

3. *Comment*

In this case there is a rather marked discrepancy between the Reputation Test results and the case worker's report on the one hand, and the impression of the mental test examiner and teacher on the other. The subsequent history of this child indicates that the opinions of the other children and of the case worker were the more accurate. Apparently Alice's record is a good example of the clarity with which even third-grade judges may diagnose a seriously maladjusted and disturbed child.

The futility of transferring an unpopular child from one school to another as a therapeutic procedure, unless it is possible at the same time to teach him better techniques, is illustrated by Alice's later career. At the end of the semester in which she was first tested, she moved to another district. After two months she was retested and had a reputation not far below average, but two years later was disliked by her new group even more cordially than by the old. However, any program for the improvement of Alice's personality would have to begin with altering that of her mother, and the prognosis for Alice's happiness is not good.

D. Case III: Margaret, a Fifth-Grade Girl

1. *Reputation Test Interpretation*

Margaret is a member of a large fifth-grade class in a middle-class school. Somewhat more than half of her classmates are boys. Girls in this room tend to choose several friends each, and in consequence a rather elaborate clique exists in which about half the girls participate either directly or via mutual friends. In this setting, Margaret names but one best friend, a popular girl who does not choose Margaret in return. She is, to be sure, mentioned once on this item, but by an isolated girl who names twelve other "friends," and it is clear that to all purposes Margaret is herself quite isolated from the others. However, she is far from inconspicuous, being mentioned by all the girls and by a larger proportion of boys than is usual at this grade level. In turn she mentions about the same number of girls as does the average fifth grader, but somewhat fewer boys. Apparently she is less aware of the personalities of her classmates than they are of hers.

The fact that Margaret receives the most negative votes from the combined group of judges of both sexes signifies that her isolation from the group's friendship associations is no accident. Both boys and girls give her the largest score on Serious (-3.0σ), thirteen girls and four boys identify her as Not Good-Looking (-2.3σ), and eleven girls and four boys mention her on Not Many Friends (-2.0σ). Girls also accord her an extreme position in the classroom on Not a Leader (-1.7σ) and Untidy (-1.1σ), and the second most extreme position on Not Good at Games, Not Friendly, and Friendly! The degree of disagreement implied on this latter pair of items (five "—" and four "+" votes) is extremely rare. Were Margaret a member of a closely knit clique, one might suspect that these scores reflect some snobbishness in her behavior, and were she simply ignored on the item, one might conclude she is not outstanding on this dimension. Since neither is the case, one may speculate that Margaret tries to be friendly but that her techniques are too poor or her behavior too inconsistent to impress all her classmates in the same way.

Despite Margaret's unpopularity she receives an extreme score on the item Quiet, and several votes on other favorable items (including Not Qarrelsome, Little Lady, Not Bossy, and Not a Show-Off), which one would scarcely expect to find in a pattern otherwise so suggestive of maladjustment. Obviously her unpopularity does not stem from over-aggressive or tomboyish behavior. However, it is difficult to account for her extreme isolation on the only grounds indicated by the test, namely, her being serious, not good-looking, and not a leader. This is a queer record in a sense, but apparently the investigator must look beyond the Reputation Test to reach more than the tentative hypothesis that Margaret is best described as serious, passive, unspontaneous, and isolated for reasons not apparent from the test record.

The teacher mentions Margaret as being Quiet and a Little Lady, but Serious and having Not Many Friends. In all four cases her opinion is in agreement with that of the class, but it fails to throw light on why a quiet, unaggressive little girl should be so unpopular with others.

Margaret's self-votes are on the items Not Quarrelsome, Takes Chances, Doesn't Fight, Bashful, and Tomboy; on none of these items is she outstanding in the opinion of the class, and so far as Tomboy, Takes Chances, and Bashful are concerned, her class reputation is opposite to her own opinion. Her self-mentions do not impress one as particularly protective, but they are certainly unrealistic if credence be given to the opinion of teacher and class.

In voting for others, Margaret distributes her votes between boys and girls and between favorable and unfavorable items much as does the average fifth-grade girl. However, she casts more votes for girls on the Popular item than does any other girl, and the second most votes for boys on the negative side of this dimension. When one considers her own reputation for not having many friends, it seems likely that her absence of

TABLE 3

Reputation Test Record of Case III, Margaret

Sex—Female	Age—10 yrs., 7 mos.	Grade—High fifth
Girls Present in Class —18.	Proportion Mentioning Margaret—100%	(Norm = 86%).
Boys Present in Class —23.	Proportion Mentioning Margaret— 65%	(Norm = 49%).
Girls Enrolled in Class—19.	Proportion Margaret Mentioned — 83%	(Norm = 84%).
Boys Enrolled in Class—24.	Proportion Margaret Mentioned — 50%	(Norm = 65%).

Item-Pairs Negative-Positive	Votes Received						Votes Given					Item Symbols
	From Girls		From Boys		Standard Score	From Teacher	To Self	To Girls		To Boys		
	Neg.	Pos.	Neg.	Pos.				Neg.	Pos.	Neg.	Pos.	
Wiggly-Quiet		*3*		**5**	.8	+						Q
Not Many Friends-Popular	**11**		*4*		−2.0	−		2	**6**	*4*		P
Serious-Full of Fun	**9**		**4**		−3.0	−		3	3			FF
Quarrelsome-Not Quarrelsome		4		1	.5		+	*4*	3		1	Ql −
Doesn't Take Chances-Takes Chances	1				− .1		+	2	**5**			C
Bossy-Not Bossy		1		1	.3							B −
Poor Sport-Good Sport	1	1		1	− .1			2	1			S
Bashful-Not Bashful		1			.1		−	1			2	Ba −
Not Good at Games-Good at Games	*6*				−1.0			1	*3*		2	G
Gets Mad Easily-Doesn't Get Mad	2			2	− .2			2	1			M −
Sissy-Real Boy	..	..	..	..	..	..	..	..	..			RB
Tomboy-Little Lady		5		3	.7	+	−	3	4	..	..	LL
Best Friend (no pair)	..	1	..		− .8			..	1	..		BF
Not Good-Looking-Good-Looking	**13**		*4*		−2.3			1	2	1		GL
Show-Off-Not a Show-Off	1	3			.2			1	2			Sh −
Not a Leader-Leader	**8**				−1.7			2	1		1	L
Untidy-Tidy	**3**				−1.1			1	2	1	4	T
Fights-Doesn't Fight		1			0.0		+			1	1	F −
Not Friendly-Friendly	*5*	*4*		2	− .2			1	3			Fr
Can't Take a Joke-Takes Joke on Self	2	1			− .4			1	1			J
Totals	62	25	12	15	..	2 − 2 +	2 − 3 +	27	38	7	11	

Note—Boldface entries indicate child was most deviate among classmates of same sex; italic entries indicate child was next to most deviate.

popularity may be a source of real tension to her. On the Serious-Full of Fun dimension, another on which her own score is low, she casts more votes than average but not the most in her class. On Takes Chances she mentions herself and casts more votes for other girls than anyone else. However, the significance of this item to the child is obscure. Her votes on Not Good-Looking and Good-Looking, the remaining dimension on which her own score is very low, are about average in number, and it is unlikely that appearance is much of an issue to her.

Summary based on reputation data. Margaret's record is very peculiar in that she is extremely unpopular in her group, but is not considered to be aggressive or to have domineering qualities usually associated with a poor reputation among girls. On the contrary, she is a follower and quiet. She is judged to be Serious, Not Good-Looking, Untidy, and Not Good at Games. Her record suggests further that she is passive, sober, and unrealistic. Apparently she has poor social techniques, but may be tense over her own unpopularity. However, she is extremely conspicuous in her group, and one can only conclude that her personality is atypical in some way not well defined by the Reputation Test results.

2. *Excerpts From the Case History*

a. Physical development. Margaret is rather large for her age and slightly above average in sex maturity.

b. Mental test results. (Stanford Form L) IQ = 151 (ten years). This has varied from 139 to 155. Examiner's comments: "Smiling and friendly—overaware of effect on people. Large and active, voice a bit unpleasant. Not bothered by failure but constantly asked, 'Is that right?' A tiring child."

c. Comments of teacher. "Daydreams all the time—always in the clouds; not friendly, diffident toward children and teacher. Not sociable, never chosen for games. Never leads and usually doesn't participate. Very quiet and somber. Not enthusiastic nor stimulated by competition. Almost too docile and courteous. Seems to try to cover up hurts. Probably sensitive, but appears thick-skinned. A social misfit." (In sixth grade was called by a different teacher "a nondescript child and very peculiar. Dreamy, reserved, toneless, and 'blah.' ")

d. From an interview with Margaret. (Obtained one month before Reputation Test.) With reference to playmates: "Oh, I always play with the whole first grade, they're so cute. Boys in my room all call me 'stink-bug.' I like them all right, but they don't like me and keep away." (Note: One other child commented that although it wasn't really true, people felt about Margaret as if she "smelled funny.") "The girls already have other friends and don't want to make new ones. They never play with me."

e. Comments of case worker. "Extremely poor social adjustment. Feels she's disliked and excluded from play. No playmates at school or in neighborhood. Lacks confidence and gets little satisfaction out of play with others. Child says she wants friends, but on the whole appears quite indifferent to people. Has practically no social relationships. Dreamy, unrealistic, and out of touch.

"According to mother, 'is very friendly, never depressed, always bubbling with enthusiasm—a rosy optimist—can amuse self for hours—has vivid imagination. Her feelings are near the surface and easily seen.' Mother detached and unrealistic—for first time admitted that Margaret has a problem in social adjustment."

3. *Comment*

Although Margaret differs markedly from Alice (Case II), the Reputation Test gives clear evidence of the social maladjustment of both. From the case histories, however, it is clear in both cases that the sources of the trouble lie in the family, and that successful psychiatric help would necessarily have to deal with the persons in the home as well as with the child herself. Margaret's case demonstrates the limitations of the Reputation Test as a diagnostic tool. Although her poor adjustment is revealed by it, nevertheless the inferences concerning her "queerness" must be based largely upon an elimination of more probable causes for rejection by other children, and cannot be very sure.

E. CASE IV: FRED, A FIFTH-GRADE BOY

1. *Reputation Test Interpretation*

Fred is a member of a fifth-grade class of thirty-five, of whom fourteen are boys. Voting on the BEST FRIEND item is heavy in this room and most boys are members of a complex clique. However, Fred is not included in this group. The only boy for whom he votes is a class leader receiving six votes, but three boys in a rather "fringe" position vote for Fred. So far as friendships are concerned, Fred's status is below average for this class.

On other items of the test he is mentioned by a very large proportion of his classmates, and in turn names more of them than the average fifth-grade boy. On most individual items, Fred's reputation score is favorable. More girls mention him than any other boy on the items QUIET, POPULAR, NOT QUARRELSOME, GOOD SPORT, NOT A SHOW-OFF, and DOESN'T FIGHT, and from them his score is second most extreme on the additional items GOOD AT GAMES, DOESN'T GET MAD, REAL BOY, TIDY, and CAN TAKE A JOKE. Boys mention him most frequently on the items QUIET, TIDY, and DOESN'T FIGHT, and second most frequently on NOT QUARRELSOME, NOT BOSSY, DOESN'T GET MAD, and FRIENDLY. However, correlational analysis (12) suggests that these traits are in general more valued by girls than by boys, and it is possible that his high status on these items may be related to his failure to be included in the clique dominated by "regular fellows." Tentative confirmation of this hypothesis is offered by several facts. First, he receives more mentions on DOESN'T TAKE CHANCES than any other boy. Second, he receives four votes on the negative and only two on the positive side of the GOOD AT GAMES dimension, although girls consider him competent. (Does he play with girls where there is less competition?) Third, he receives as many votes from boys on SISSY as on REAL BOY, although characterizations on the latter item are more common. This pattern of reputation scores suggests that Fred is unaggressive and friendly, but is held in higher regard by girls than by boys, and is mentioned for qualities which are more appropriate in girls than in boys, especially QUIET (3.0σ), NOT QUARRELSOME (2.1σ) DOESN'T FIGHT (2.0σ), and DOESN'T TAKES CHANCES (-1.0σ).

Fred's teacher names him on NOT BOSSY, NOT A SHOW-OFF, DOESN'T FIGHT, and FRIENDLY, but mentions that he DOESN'T TAKE CHANCES and is a POOR SPORT.

Fred's record of self-mentions is unique among all the children to whom the test was given, in that he names himself on 15 of the 18 favorable items. His only negative self-nomination is DOESN'T TAKE CHANCES, but equally unusual by comparison with the average boy are his self-mentions on QUIET and GOOD-LOOKING. Item-pairs omitted by Fred are NOT BASHFUL–BASHFUL and LEADER–NOT A LEADER, on both of which he receives a small negative score from his classmates. Since Fred's reputation with others is on

TABLE 4

REPUTATION TEST RECORD OF CASE IV, FRED

Sex—Male	Age—10 yrs., 2 mos.	Grade—Low fifth
Girls Present in Class —21.	Proportion Mentioning Fred— 86%	(Norm = 64%)
Boys Present in Class —13.	Proportion Mentioning Fred—100%	(Norm = 79%)
Girls Enrolled in Class—21.	Proportion Fred Mentioned — 62%	(Norm = 49%)
Boys Enrolled in Class—14.	Proportion Fred Mentioned — 92%	(Norm = 79%)

Item-Pairs Negative-Positive	Votes Received						Votes Given					Item Symbols
	From Girls		From Boys		Standard Score	From Teacher	To Self	To Girls		To Boys		
	Neg.	Pos.	Neg.	Pos.				Neg.	Pos.	Neg.	Pos.	
WIGGLY-QUIET		**10**		**5**	3.0		+	1	*3*	3	1	Q
NOT MANY FRIENDS-POPULAR		**3**		5	1.2		+		*2*	**5**	2	P
SERIOUS-FULL OF FUN			1	*3*	.3		+	1		2	3	FF
QUARRELSOME-NOT QUARRELSOME		**5**		*6*	2.1		+		1	2	1	Ql —
DOESN'T TAKE CHANCES-TAKES CHANCES	1		**3**		−1.0	—	—			1	1	C
BOSSY-NOT BOSSY		1		*4*	1.1	+	+	1	*1*	2	1	B —
POOR SPORT-GOOD SPORT		**5**	1	2	1.1	—	+			1	2	S
BASHFUL-NOT BASHFUL				1	− .1				1	1	**3**	Ba —
NOT GOOD AT GAMES-GOOD AT GAMES		*2*	4	2	− .3		+	**3**	**4**	1	3	G
GETS MAD EASILY-DOESN'T GET MAD		*2*		*4*	1.3		+			2	2	M —
SISSY-REAL BOY	2	*9*	*3*	3	.5		+	..	..	2	3	RB
TOMBOY-LITTLE LADY	..	..	..	..	..	..	..	1	3	..	..	LL
BEST FRIEND (no pair)	..		..	3	.5			..		..	1	BF
NOT GOOD-LOOKING-GOOD-LOOKING		*3*		1	1.0		+		1	**4**		GL
SHOW-OFF-NOT A SHOW-OFF	1	**3**		4	1.4	+	+			1	1	Sh —
NOT A LEADER-LEADER		1	2		− .3				1	1	1	L
UNTIDY-TIDY		*2*		**4**	1.3		+		2	2		T
FIGHTS-DOESN'T FIGHT		**4**		**7**	2.0	+	+			1	1	F —
NOT FRIENDLY-FRIENDLY		1		*5*	1.7	+	+	**3**	2		2	Fr
CAN'T TAKE A JOKE-TAKES JOKE ON SELF		*2*		3	1.4		+			1	2	J
Totals	4	53	14	62	..	2 — 4 +	1 — 15 +	10	21	32	30	

Note—Boldface entries indicate child was most deviate among classmates of same sex; italic entries indicate child was next to most deviate.

the whole favorable, it is likely that this very unusual degree of self-approbation is more indicative of complacency than of compensatory tendencies. One wonders if Fred would not have reversed his lone negative self-vote on DOESN'T TAKE CHANCES, had he realized that by the statistical conventions of the tabulation it would receive a minus score.

Fred casts more votes for others than average, and makes relatively more mentions of girls than of boys as compared with the norm, suggesting again that he has more contact with, or interest in them. He mentions more girls on both poles of the GOOD AT GAMES and FRIENDLY dimensions than any other boy. He casts more votes for boys than any of his classmates on the items NOT MANY FRIENDS, NOT BASHFUL, and NOT GOOD-LOOKING, but his only nominations for boys on GOOD-LOOKING and TIDY are for himself. This suggests that he is smug about himself but unusually critical of other boys. However, the traits on which he considers himself superior are regarded by most boys as more appropriate to girls.

Summary based on reputation data. Fred's record indicates that he is a "good boy,"—almost too good to be quite credible. He receives many favorable nominations from children of both sexes, but they tend to be on items which are more important for the adjustment of girls than of boys, and in actual friendship associations Fred is rather excluded from the group. He is friendly and unaggressive, but apparently quite complacent and conceited. He is probably cautious, not very good at games, and not quite a "real boy," his own opinion to the contrary. He seems to have relatively more contact with girls than do most fifth-grade boys, and is better liked by girls than boys. Fred's apparent personality pattern comes close to the children's conception of an ideal girl, but one has some fears for his social adjustment as he enters adolescence where more masculine behavior is demanded of him.

2. *Excerpts From the Case History*

a. Physical development. Fred is above the norm in size, but slower to mature than average.

b. Mental test results. (Stanford Form M) IQ = 149. (Has varied from 125 to 156.) Examiner's comments: "Well poised and friendly. Makes effort to do well—very earnest and tries hard to please."

c. Comments of teacher. (Data obtained six months after Reputation Test but from same teacher.) "Extremely popular, pliable, courteous child. Lovable and likes others. Excellent behavior and attitude toward school. Good sense of humor—never shows off. Above average academic performance. Is product of a wonderful home environment."

d. From an interview with Fred. "All the boys I know like me—I have lots of friends. Dick is my best friend. He's swell and I'm good, too. He and I are about the best in school."

e. Case worker's comments. "Likes teacher, likes school. Is always cooperative in interview, but cautious until assured of a sympathetic audience. Is rather quiet—not a show-off. Always good-humored and even-tempered. Dislikes getting into arguments. Extremely sympathetic and generous. Perhaps a little dependent on mother, who is determined not to have Fred a 'sissy' but scolds him if he comes home after play with his clothes torn or dirty. Seems to put forth a great effort to do the correct or approved thing in all situations. Is self-satisfied and takes himself seriously in his attempt to live up to his ideal of perfection. Is wholly unaware of his own liabilities, and receives enough success at school and at home to convince him that he is succeeding in reaching his goal." (Two years later the comment was made: "Role of being the perfect boy seems to have become so habitual that it no longer requires much strain.")

3. *Comment*

Fred reveals himself on the Reputation Test to a surprising degree, and the interpretation coincides with other data concerning him. His favorable reputation on many items viewed in relation to his exclusion from the clique organization among his classmates is more evidence of the truism that exemplary characters are more apt to be respected than sought out as friends. Noteworthy is the fact that at adolescence Fred lost status and was almost isolated by the other boys, whose group values at this stage were for active, aggressive, nonconformist qualities which Fred lacked. Despite the over-all favorable impression of his teacher, the Reputation Test correctly forecast the nature of these adjustive difficulties in which Fred later found himself.

F. Case V: James, a Fifth-Grade Boy

1. *Reputation Test Interpretation*

James is a member of a fifth-grade class containing almost twice as many girls as boys. More than half of the children enjoy reciprocal friendships with other members of the class. Although there are four "stars" of exceptional popularity, no complex clique groupings can be discerned. In this group James chooses as Best Friend a boy who in turn chooses him, but neither receives any votes from anyone else. James's mention as a Best Friend of a girl (who, however, doesn't mention him) is sufficiently uncommon to deserve mention. On the test as a whole, James is mentioned by fewer boys and by fewer girls than is the average fifth-grade boy, but he mentions considerably more than the average.

The votes James receives on negative items, both from boys and from girls, are almost twice as numerous as are his favorable mentions. However, since he is relatively inconspicuous, the total number of unfavorable judgments is not large, and on no item-pair is the boy's standard score larger than —.8σ. His more deviate scores are received on the items Not a Leader (—.8σ), Sissy (—.8σ, although three girls considered him a Real Boy), Best Friend (—.7σ), Doesn't Take Chances (—.6σ), Not Good at Games (—.5σ), and Serious (—.5σ).

None of these items is correlated with Real Boy and its associated variables. The implication is plain that James is, if not somewhat effeminate, at least not as masculine as the average. This impression is substantiated by the fact that the items on which he received a favorable standard score (Quiet, Not Quarrelsome, Not Bashful, Tidy, and Doesn't Fight) are traits more valued by girls than by boys.

Jame's teacher mentions him as Popular, Full of Fun, A Good Sport, A Real Boy, Friendly, and most significant, as her Best Friend, an item usually omitted by teachers. This wholly favorable picture is quite at variance with that drawn by his classmates. It suggests that James gets along better with adults than with children, and probably presents a pretty bland front to the former. In this case it may be that the boy is suffering mild hostility from his classmates as a result of being too well liked by his teacher. It is further possible that the approval he receives from his teacher indicates that he is brighter academically than most of his group.

Despite his favorable and secure position from the point of view of his teacher, some evidence of insecurity is offered by the fact that he names himself only on unfavorable qualities, and makes the next to the most such mentions of any boy in class. (The average boy is more prone to name himself for favorable qualities.) Some of

TABLE 5

REPUTATION TEST RECORD OF CASE V, JAMES

Sex—Male	Age 11 yrs., 1 mo.	Grade—High fifth
Girls Present in Class —25.	Proportion Mentioning James— 52%	(Norm = 64%).
Boys Present in Class —14.	Proportion Mentioning James— 69%	(Norm = 79%).
Girls Enrolled in Class—25.	Proportion James Mentioned — 72%	(Norm = 49%).
Boys Enrolled in Class—14.	Proportion James Mentioned —100%	(Norm = 79%).

Item-Pairs Negative-Positive	Votes Received						Votes Given					Item Symbols
	From Girls		From Boys		Standard Score	From Teacher	To Self	To Girls		To Boys		
	Neg.	Pos.	Neg.	Pos.				Neg.	Pos.	Neg.	Pos.	
WIGGLY–QUIET					.3			1	2	1	1	Q
NOT MANY FRIENDS–POPULAR				1	0.0	+		*3*		1	3	P
SERIOUS–FULL OF FUN	1				− .5	+		1	*2*	2	2	FF
QUARRELSOME–NOT QUARRELSOME			1	3	.4			1	*4*	2	1	Ql −
DOESN'T TAKE CHANCES–TAKES CHANCES			2		− .6		−			2	1	C
BOSSY–NOT BOSSY	1				− .1			4	4	1	1	B −
POOR SPORT–GOOD SPORT				1	0.0	+		1	3	3	3	S
BASHFUL–NOT BASHFUL				2	.1			1				Ba −
NOT GOOD AT GAMES–GOOD AT GAMES		1	3		− .5		−	*2*	4	1		G
GETS MAD EASILY–DOESN'T GET MAD		2	2		0.0		−	1	4	3		M −
SISSY–REAL BOY	4	3	3		− .8	+		..	..	1	6	RB
TOMBOY–LITTLE LADY	..	..	..	..	..	..	..	1		..	..	LL
BEST OF FRIEND (no pair)	..		..	1	− .7	+		..	1	..	1	BF
NOT GOOD-LOOKING–GOOD-LOOKING	2				− .4							GL
SHOW-OFF–NOT A SHOW-OFF	3		1	1	− .3				5	1		Sh −
NOT A LEADER–LEADER			4		− .8		−					L
UNTIDY–TIDY			1	1	.1							T
FIGHTS–DOESN'T FIGHT					.1		−			1		F −
NOT FRIENDLY–FRIENDLY					− .1	+						Fr
CAN'T TAKE A JOKE–TAKES JOKE ON SELF			1	1	− .3			1	3			J
Totals	11	6	18	11	..	0− 6+	*5*− 0+	17	32	19	19	

Note—Boldface entries indicate child was most deviate among classmates of same sex; italic entries indicate child was next to most deviate.

his self-mentions, viz., DOESN'T TAKE CHANCES, NOT GOOD AT GAMES, and NOT A LEADER, indicate his awareness of being less venturesome and ascendant than other boys. His other self-votes (GETS MAD and FIGHTS) are not in harmony with the opinion of the class, and may indicate aggressive wishes or a fantasy of being more aggressive than he really is. In any event, he seems franker and better able to verbalize his problem than most boys in similar circumstances.

James is relatively more interested in or at least more aware of the personalities of girls than of boys. He casts more votes for girls than any other boy on seven items, and the next to the most on four more, although on no item is he correspondingly extreme for mentions made of boys. Not only does he cast more votes for girls than for boys, but relatively more of them are favorable. He casts the most votes of any boy for girls both on BOSSY and on NOT BOSSY, and makes a similarly extreme number of mentions of girls on the GOOD AT GAMES–NOT GOOD AT GAMES pair. On the other hand, in mentioning boys, he names only one as BOSSY and one as NOT BOSSY, omits GOOD AT GAMES (usually an item eliciting heavier voting than any other in the test), and names only one boy as NOT GOOD AT GAMES. These facts taken together suggest quite strongly that he does not play much with boys but rather enters, or at least tries to enter, into the girls' games where he is sometimes rebuffed by the "bossier ones." However, the fact that he makes six nominations on the REAL BOY item indicates that this is an important dimension for him and possibly one concerning which he feels sufficiently uneasy to omit a self-mention on either pole.

Summary based on reputation data. James is probably a bright, polite boy, more attractive to adults than to other children, though not actually unpopular with anyone. He is not very masculine, is a follower, cautious, unaggressive, and lacking in athletic skills. He is probably open and outgoing but quite insecure, and may harbor aggressive wishes which he is unable to carry into action. He seems to know more about girls than boys, and to esteem them more highly. However, he is not too well thought of by members of either sex. Fortunately, he has one close friend in the class who may give him considerable support. He may be headed for an unhappy time in adolescence. Efforts to improve his physical skills might well be expended on his behalf.

2. *Excerpts From the Case History*

a. Physical development. James is slightly above average in height, but very obese (girdle adiposity). Sex maturation definitely below average. Has had a kidney disorder of long standing, and in consequence his physical activity has been curtailed on medical advice.

b. Mental test results. (Stanford, Form L) IQ = 149. (This has varied from 137 to 149.) Comment of Examiner: "Interested in success but not upset by failure. Friendly, relaxed, and gracious. Cooperative and kindly. Good insight—a few show-off mannerisms, but quite well poised."

c. Teacher's comments. "Very friendly and well liked, witty and humorous—self-reliant, quite courteous. A leader admired by others, has ideas they accept. Enjoys excelling, and does. Grades good or excellent. Doesn't join in hard games because of physical condition. Quite a thinker and reader—can converse on adult level. A well-adjusted child."

d. From an interview with James. With reference to social adjustment: "I used to be shy but I have lots of friends now. I always arrange to meet Robert (named as friend on the Reputation Test) at recess." (Q) "We don't play much—just walk around."

With reference to teasing: "Sometimes kids say, 'Hi-ya, fatty?' but I don't let it get under my skin unless they do it too much."

With reference to scholastic work: "I always get a good card but not better than most." (Note: One year later James commented that he knew lots of people but didn't have bosom friends. He admitted that he was not popular, but he thought boys like him better than girls.)

e. Case worker's comments. "Is overdependent upon mother. Demands attention and is lazy. Very sensitive to remarks about his size. Too cautious, never wants to try anything new. Is friendly, expressive, and spontaneous. Dislikes quarrels and keeps out of them. Good-natured—gets along well with adults and children. Prefers to win, but isn't too disturbed by failure. High energy level, a verbal child with an adult vocabulary. Seems to be quite content with his lot in life."

3. *Comment*

James's major problem, apart from his physical handicap, seems to lie in the field of social adjustment to his contemporaries. It seems obvious that his difficulties are related to his inability to participate in active sports, but a solution to them is difficult. It is interesting that the Reputation Test revealed James's rather poor status at a time when adults were giving this polite and verbal child a much more favorable characterization, and a year before the boy himself admitted the true state of affairs.

G. CONCLUSION

Illustrative clinical interpretations have been presented based on the Reputation Test records of five Guidance Study subjects chosen to represent some of the more deviate patterns obtained. Teachers' mentions, self-mentions, reputation *with* and votes given *to* other classmates were considered in relation to the norm for the appropriate sex and grade level, and to the pattern of friendships revealed by the BEST FRIEND item.

The adequacy of a clinical characterization of a child based on his reputation record depends, apart from the insightfulness of the interpreter, upon how deviate the child is in his group and upon how full his record is. However, an extremely scant number of votes given and received may itself be indicative of a withdrawn personality. Behavior which is not correlated with that elicited in the school situation cannot be predicted on the basis of the Reputation Test, and nonsocial aspects of personality are also difficult to get at. Statements about more obvious personality characteristics can be made quite accurately, within the limitations noted above, on the basis of the reputation data taken alone. Inferences about a child's areas of tension derive from his reputation, and in addition from his pattern of voting (i.e., from the projective aspect of the test), but the significance of the latter can be evaluated only in relation to the former.

We conclude that although the Reputation Test is limited both as to the areas of personality which it can reveal and as to the types of questions which it is expedient to include in a classroom setting, it is nevertheless useful in diagnosing a child's adjustment to his classmates, and can throw some light on his social maturity and on his areas of sensitivity and conflict. Since the Reputation Test yields information that is not always directly accessible to teachers making judgments from an adult point of view, it frequently reveals problems making for social maladjustment much earlier than they are ordinarily detected by adult observers. Hence, it may make a practical contribution to counselors and school administrators, both in locating children who are having psychological difficulties in the field of social relationships, and in measuring the effectiveness of efforts on their behalf.

REFERENCES

1. Campbell, Helen M. Sex differences obtained by the "guess who" technique in reputation assessments given and received by adolescent boys and girls. Unpublished Doctor's thesis. Univer. Calif., 1940.
2. Elliott, M. H. Patterns of friendship in the classroom. *Prog. Educ.*, 1941, 18, 383-390.
3. Hartshorne, H., May, M., & Maller, J. B. *Studies in the nature of character: II. Studies in service and self-control.* New York: Macmillan, 1929.
4. Jones, H. E. The adolescent growth study. *J. consult. Psychol.*, 1939, 3, 157-159 and 177-180.
5. Macfarlane, Jean W. Studies in child guidance. I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, 3 (no. 6).
6. Macfarlane, J. W., Honzik, M. P., & Davis, M. H. Reputation differences among young school children. *J. educ. Psychol.*, 1937, 28, 161-175.
7. Moreno, J. L. Who shall survive? A new approach to the problem of human interrelations. *Nerv. ment. Dis. Monogr. Ser.*, 1934, no. 58.
8. Symonds, P. M., & Jackson, C. E. *Measurement of the personality adjustments of high school pupils.* New York: Columbia Univer. Press, 1935.
9. Tryon, Carolyn M. A study of social and emotional adjustments. Unpublished Doctor's thesis, Univer. Calif., 1933.
10. Tryon, Carolyn M. Evaluations of adolescent personality by adolescents. *Monogr. Soc. Res. Child Develpm.*, 1939, 4 (no. 4).
11. Tuddenham, R. D. Belonging in a group. *Educ. Leadership*, 1944, 1, 201-205.
12. Tuddenham, R. D. Studies in reputation: III. Correlates of popularity among elementary school children. *J. educ. Psychol.*, 1951, 42, 257-276.
13. Tuddenham, R. D. Studies in reputation: IV. Patterns in friendship among school children. Unpublished ms., Univer. of Calif. Inst. Child Welfare, 1950.

VOLUME 66 WHOLE NO. 334
NUMBER 2 1952

Psychological Monographs: General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Reliability and Validity of Judges' Ratings of Adjustment on the Rorschach

By

MARGUERITE Q. GRANT, VIRGINIA IVES,
and JANE H. RANZONI
Institute of Child Welfare, University of California

Accepted for publication March 28, 1951

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

I. INTRODUCTION[1]

THIS research was designed to answer three main questions which grew out of previous Rorschach research in the Guidance Study at the Institute of Child Welfare, University of California. These questions are: (1) Can Rorschach "experts" agree on ratings of adjustment based on the Rorschach? (2) In what way do increasing amounts of Rorschach material affect such judgments? (3) If substantial inter-rater agreement can be demonstrated, what relationship can be shown with outside criteria?

In a previous article from the Institute of Child Welfare (5), the possibilities and limitations inherent in the "sign" approach to the Rorschachs of a normal adolescent population have been explored. The article presented data indicating that the incidence of the so-called "neurotic" signs in this population was so high as to make meaningless the label "neurotic." The possible hypothesis that adolescence is a time of emotional turmoil and that one could, therefore, expect many neurotic signs, was not supported by the data. With this hypothesis one would predict a decrease in the number of signs as the crisis period in adolescence began to merge into adulthood. In our group, no consistent decline was evident over the age range 11 to 18 years. Conversely, the indicators of normality (Rorschach "adjustment" signs) were low; their correlations with interview ratings of adjustment varied nonsignificantly around zero.

The results of that study made it apparent that the "sign" method offered little in a systematic and fruitful approach to the understanding of the Rorschachs of our cross section of adolescents. These findings may not be surprising to many Rorschach workers. One of the major characteristics claimed for the test has been that, for a meaningful understanding of the dynamics of personality organization, the Rorschach must be used as a gestalt. Clinicians agree that the more atomistic the Rorschach analysis, the less its worth as an evaluative instrument. With this in mind, for the present study, three experienced Rorschach workers were asked to make subjective judgments of adjustment on the basis of the Rorschach records.

II. THE SAMPLE AND CRITERION

The subjects were 71 boys and 75 girls who were members of a longitudinal growth study and had been randomly selected 18 years previously from the birth registry in Berkeley, California.[2] The Rorschach had been administered as part of the testing program of the study, beginning when the children were 11 years of age. As a result, the 18-year records used in this experiment are the fourth Rorschach for most of the subjects. These records were chosen for three reasons. First, they are the most "adult" records and, therefore, more comparable to other "adult" norms in the literature. Second, age 18 has the largest number of cases at any one age in this sample, and, third, these records are longer and richer than earlier tests and thus offer a better basis for rating.

---

[1] Clerical work and statistical analyses were made available by a grant-in-aid from the U. S. Public Health Service. The study proper was made possible by assistance from the Laura Spelman Fund, the General Education Board, and the Rockefeller Foundation.

Special acknowledgement is due to Jean W. Macfarlane, Director of the Guidance Study, Institute of Child Welfare, for her help and suggestions throughout this study; to Lucile Allen for making special adjustment ratings of the C group; to Helen Clementson for acting as one of the judges; to Edith Katten and Arline Burton for their roles as control raters; and to Frances G. Orr and Margaret Snyder for their work in administering the Rorschachs. The authors also wish to thank Dr. Rheem Jarrett and Charles Hanley for their statistical advice, and Brittomar Handlon for assistance in computation.

[2] For a complete description of the sample, see Macfarlane (8). For a discussion of Rorschach procedures in the Guidance Study at the Institute, see McFate and Orr (9).

The outside criteria of adjustment consist of ratings of current adjustment made by case workers and based on considerable interview contact with the subject over a period of years. The areas of self-adjustment, social adjustment, and family adjustment were among the variables rated. These ratings and the interviews upon which they were based will be described in detail in the section on validity.

III. PROCEDURE

A. JUDGES

The ratings of the Rorschach protocols were made by three women judges. Since ratings depend upon the training and Rorschach experience of raters and the breadth of the sample upon which the experience is based, a brief statement about each rater follows.

Judge A was trained by Bruno Klopfer in 1939 and has subsequently attended several "workshops." In the course of ten years of administering and teaching the Rorschach, she has seen and evaluated well over 5,000 records, including all ages and both normal and pathological cases.

Judge B was initially trained by Judge A in 1943 and has subsequently had several "workshops" with Bruno Klopfer. She has worked with other psychologists, well trained in the Rorschach, both in the clinic and in academic settings, and has spent two years teaching and supervising Rorschach work in a clinic.

Judge C attended a Hertz "workshop" in 1946 and has since used the test in a well-established clinic under expert supervision.

For purposes of control, two other raters were used in part of the experiment. The Rorschach experience of these "nonexpert" raters is described in Section IV, Part 4.

B. THE RORSCHACH RATING SCALE

A four-point scale was used for sorting the Rorschach: (I) very maladjusted, (II) moderately maladjusted, (III) fairly well adjusted, and (IV) very well adjusted. These scale points were not further defined.

C. THE SORTING PLAN

The records were divided into four groups for the purpose of rating: 41 Guidance (*G*) group boys; 37 Guidance (*G*) group girls; 30 Control (*C*) group boys; and 38 Control (*C*) group girls.[3] Three sorts into the four adjustment categories were made on each group of records, with successively increasing amounts of Rorschach information given to the judges on each. This plan was followed in order to determine the contributions of both the formal aspects and the content to the reliability of the ratings and the capacity of the tool to predict clinical judgments.

For Sort 1, the only Rorschach material on each case consisted of the summary page of the standard Klopfer and Davidson scoring form, including the determinant profile and the standard percentages and ratios. In addition, the raters were given the norms of this particular 18-year-old sample, in the form of means, medians, quartiles, and per cent occurrence for each of the four groups.

For Sort 2, the judges were given the previously mentioned summary page plus the tabulation page of the scoring form, which shows the distribution of the scored variables over the ten cards. On Sort 3, the raters were given the entire record, including the full protocol.

---

[3] The primary difference between the *G* (Guidance) and *C* (Control) groups is the intensity with which they have been studied through interviews. The Rorschach administrations were equivalent for the two groups.

The protocols were presented in a random order, with this order changed from sort to sort. Any case-identifying material was deleted, since one of the judges had also acted as an interviewer with some of the children whose Rorschachs were used, but the sex of the subject was known. On each sort, both boys and girls in the *G* or *C* groups were rated before continuing to the next sort, so that one to two days elapsed between the rating and rerating of the same Rorschach records. The judges did not have access to their previous ratings. This fact, plus the size of the groups sorted, probably minimized the memory factor in serial rating of the same records.

The instructions to the judges were (*a*) to take into consideration both self- and social adjustment in their ratings, (*b*) to strive for a normal distribution of ratings over the four-point scale, (*c*) to try to rate adjustment in terms of the norms of *this* group, and (*d*) to record separately the cues used for individual ratings when this was possible.

After the three sorts were completed on all groups, the judges conferred in an attempt to make explicit the cues they had used in their ratings and discussed specific cases in which disagreement occurred.

IV. RESULTS AND DISCUSSION

A. Rater Behavior

1. *Distributions of ratings over three sorts*

All of the judges found it difficult to approximate a normal distribution on the four-point adjustment scale even though they had been asked to attempt it and to rate in terms of this group. The distributions of the cases remained quite constant as more Rorschach information was given to the raters. (See Table 1. The distribution of cases was so similar for all judges that the results of the three judges' ratings are averaged for this table.) On all three sorts for both boys and girls, 61 to 71 per cent of the cases were rated as "maladjusted" (Ratings I and II). With the exception of the first sort of the boys' group, the differences in the per cent of cases rated as "maladjusted" and those rated as "adjusted"

TABLE 1

Percentage Distribution of Ratings over Three Sorts

Adjustment Category	Boys ($N=71$)*			Adjustment Category	Girls ($N=75$)*		
	Sort				Sort		
	1	2	3		1	2	3
I	21%	27%	30%	I	16%	16%	17%
II	40%	44%	41%	II	50%	49%	50%
III	31%	25%	27%	III	26%	30%	26%
IV	8%	4%	2%	IV	8%	5%	7%
	100%	100%	100%		100%	100%	100%

Sort 1—Profile page only
Sort 2—Profile page plus tabulation sheet
Sort 3—Entire record including content

Adj. category I—very maladjusted
Adj. category II—moderately maladjusted
Adj. category III—fairly well adjusted
Adj. category IV—very well adjusted

* Sorts were done in four groups: *G* boys, $N=41$; *C* boys, $N=30$; *G* girls, $N=37$; *C* girls, $N=38$.

are statistically significant above the 5 per cent level of confidence. Many more cases were rated as "very maladjusted" than were judged "very well adjusted" by all three raters. A larger percentage of boys than of girls were rated "very maladjusted" by all judges. There is some tendency for the percentage of boys placed in adjustment Category I to increase as more Rorschach information is given the judges. A corresponding decrease occurs in the number of cases placed in Category IV. This change does not take place in the girls' groups.

2. *Intra-rater agreement*

As more Rorschach material was made available over the three sorts, the judges tended to change the adjustment rating of approximately half of the cases by at least one category (see Table 2), in spite of the fact that the percentage of cases in each adjustment group remained quite constant. It might have been expected that the greatest number of changes in rating would occur between Sorts 1 and 3. This was true in the case of the girls' groups, but not consistently in the boys'. Other sex differences are also evident. In the ratings of all judges, the least change for the girls' groups occurs between Sorts 1 and 2, but for the boys' groups, between 2 and 3. These sex differences seem in line with many other findings in the Guidance Study Rorschach studies. That is, the content and verbalizations (introduced in Sort 3) of the girls' records are much richer than those of the boys', and so might tend to influence to a greater extent the change in the Rorschach adjustment ratings given to the girls.

3. *Inter-rater agreement*

Table 3 presents an affirmative answer to one of the questions raised in the introduction, namely, whether experts can agree on over-all adjustment ratings made from Rorschach protocols. Although far from unity, the tetrachoric correlations of inter-rater agreement all show an acceptable degree of relationship between the ratings of the three judges.

The sex difference is again apparent. In rating the boys' records, there is an increase in inter-judge agreement from the first to the second sort and a slight decrease from the second to the third sort as more Rorschach information is given. The correlations decrease slightly on the second sort for the girls, then increase on the third.

A comparison of the full and partial

TABLE 2

INTRA-RATER AGREEMENT:
PERCENTAGE OF JUDGES' SELF-AGREEMENT IN RATINGS OVER THREE SORTS

Per Cent of Cases Constant Between:	Boys ($N=71$)			Per Cent of Cases Constant Between:	Girls ($N=75$)		
	Rater				Rater		
	A	B	C		A	B	C
Sorts 1 and 2	61%	51%	54%	Sorts 1 and 2	73%	61%	60%
Sorts 2 and 3	63%	59%	61%	Sorts 2 and 3	61%	48%	57%
Sorts 1 and 3	56%	54%	54%	Sorts 1 and 3	45%	43%	49%

Sort 1—Profile page only
Sort 2—Profile page plus tabulation sheet
Sort 3—Entire record including content

TABLE 3

CORRELATIONS OF INTER-RATER AGREEMENT ON ADJUSTMENT RATINGS*

Raters	Boys ($N=71$)			Raters	Girls ($N=75$)		
	Sort				Sort		
	1	2	3		1	2	3
A×B	.61	.85	.68	A×B	.70	.65	.72
A×C	.67	.71	.70	A×C	.45	.42	.66
B×C	.61	.75	.72	B×C	.73	.50	.60

* Significance of the tetrachoric correlations is indicated as follows:
Significantly different from zero at 1 per cent level, ═══
Significantly different from zero at 5 per cent level, ———

agreement of the judges' ratings is shown in Table 4. (Since all of the distributions were so skewed, these are probably a better measure than the tetrachoric correlations.) Over the three sorts, on 29 per cent of the girls and 34 to 39 per cent of the boys, the three judges agreed completely on the adjustment rating given. *When, to the cases of complete agreement, we add those cases in which two judges agreed on a rating category and the third judge disagreed by one scale point only, 85 to 98 per cent of all cases are included.* The other 2 to 15 per cent consist of those cases in which two judges agreed and the third judge disagreed by two scale points or in which the three judges gave three different ratings. In addition to the above-mentioned lower percentage of girls than boys on which the judges completely agreed, Table 4 shows a higher percentage of girls on which the judges completely disagreed.

TABLE 4

INTER-RATER AGREEMENT AND DISAGREEMENT ON THREE SORTS

Extent of Agreement	Boys ($N=71$)			Girls ($N=75$)		
	Sort			Sort		
	1	2	3	1	2	3
Three judges agree	39%	34%	38%	29%	29%	29%
Two judges agree and third judge disagrees by one scale point	46%	64%	58%	58%	62%	56%
Two judges agree and third judge disagrees by two scale points	8%	1%	0%	4%	4%	3%
Two judges agree and third judge disagrees by three scale points	0%	0%	0%	0%	0%	0%
Three judges disagree	7%	1%	4%	9%	5%	12%
	100%	100%	100%	100%	100%	100%

Sort 1—Profile page only
Sort 2—Profile page plus tabulation sheet
Sort 3—Entire record including content

It was expected that because of similarity of training and experience background, Judges A and B would agree more highly with each other than either would with Judge C. Neither the evidence in Table 3 nor a further breakdown of Table 4 into the agreements and disagreements of individual raters indicates any such tendency. The average per cent agreement of the judges over all sorts is as follows: A with B, 56 per cent; A with C, 51 per cent; B with C, 53 per cent.

4. *Sorting by Rorschach "novices"*

After discovering the extent to which Rorschach "experts" could agree on adjustment ratings from Rorschach protocols, the authors asked two more individuals to act as judges in rating one of the four groups described in this paper. One of them, Judge D, has had many years of experience interviewing as a social worker. With respect to the Rorschach, she can be considered "naive," since her only experience with the test was as a subject a year prior to this study. Judge E is a graduate student in clinical psychology, who had, at the time of these ratings, taken a one-semester course in Rorschach administration and scoring from Judge A. Since the formal scoring aspects of the records meant nothing to Judge D and relatively little to Judge E, they were given the entire protocols to rate. Thus, their one sort was comparable to Sort 3 for Judges A, B, and C. Judge D worked with the records of the *C* girls ($N = 38$) and Judge E with the records of the *G* boys ($N = 41$).

Table 5 presents the inter-rater agreement of the Rorschach "novices" with the "experts" and the agreement among the "experts" in the ratings of the groups with which the "novices" worked. Comparing these results, we find that Judge D's correlations with the original judges are only slightly lower than the "experts" agreements among themselves and that Judge E correlates higher with Judges A, B, and C than they did with each other.

TABLE 5
INTER-RATER AGREEMENT (TETRACHORIC r's) OF RORSCHACH "NOVICES" WITH "EXPERTS," AND OF "EXPERTS" WITH EACH OTHER, ON COMPARABLE SORTS*

Raters	Rater D	Raters	Experts**
	C Girls		*C* Girls
D×A	.55	A×B	.63
D×B	.33	A×C	.55
D×C	.49	B×C	.57
Raters	Rater E	Raters	Experts
	G Boys		*G* Boys
E×A	.70	A×B	.70
E×B	.85	A×C	.61
E×C	.75	B×C	.73

* Significance of the tetrachoric correlations is indicated as follows:
Significantly different from zero at 1 per cent level, ═══
Significantly different from zero at 5 per cent level, ———

** In order to simplify the data, *G* and *C* groups were combined for the inter-rater agreement correlations presented in Table 3. Consequently, the correlations presented here do not appear in Table 3.

Does the magnitude of the inter-rater agreement among the "experts" mean that the Rorschach protocol as a behavior sample contains communicable information concerning the subject, or does it simply mean that through intensive and similar training, the three judges have learned to a degree to *do* the same things when confronted with the test material? The former may be the case, since judges who have had little or no training are able to make similar translations of Rorschach records. Apparently several months of training and years of experience in the application of a complicated scoring system have contributed little to the agreement among the "experts" when the discriminations called for are no finer than a four-point scale.

5. *Subjective impressions of the judges*

Recorded conferences of the "experts" following their ratings made it clear that, although the raters individually were concerned with different aspects of the problem, all three agreed that it was very difficult to make over-all adjustment ratings from these Rorschachs. They felt little confidence in many of their ratings. They continually reviewed their judgments in an attempt to achieve a more normal distribution on the rating scale, yet were unable to do so. They found it difficult to rate the adjustment of these cases in terms of the norms of this group, which were available to them, because it was hard for them to depart from their preconceived notions of a "normal" Rorschach. For example, standard assumptions of normal Rorschach balance include greater use of color and a lower *F%* and *A%* than are found in this sample.

The judges agreed that the boys' groups presented quite different sorting problems from the girls' groups. The girls' records were much richer with regard to both formal aspects and content, and the judges felt that they offered much more basis for the adjustment rating. The greater poverty of the boys' records, together with the judges' impressions of general "ugliness" of content, made it difficult for the raters to give them ratings of III or IV. In fact, on the final sort of one of the boys' groups, each judge independently failed to place any record in the "well-adjusted" category.

On the above problems, the judges were in substantial agreement. In addition, each of the raters raised particular issues which were of individual concern during the sorting process. Judge B felt it was necessary to distinguish between present adjustment and predicted future adjustment, since in many cases she would give different ratings. For instance, a girl of 18 whose Rorschach shows a tendency to relate only on a superficial level might be reasonably well adjusted in her 18-year-old peer culture, but might encounter difficulties with emotional relationships later. Judge A wanted to emphasize the necessity of different adjustment criteria for boys and girls in terms of the different cultural patterns demanded of the two sexes. For example, while fewer *W* than *D* and a *W:M* ratio of less that 2:1 might be negative signs in a boy's record, they would not necessarily be so in a girl's test.

Many of the cues for ratings which the judges verbalized were those which have been discussed often in the literature and which have been utilized particularly in the "sign" lists of Davidson, Monroe, etc. However, the judges used them not as isolated items but as parts of a configuration which determined their weights. A few examples of other cues on which the judges were in essential agreement follow. Overemphasis on any particular blot aspect (too many *W, M, Fc,* etc., responses in proportion to the rest of the record) was considered a negative sign. The use of achromatic color or white space in a colored card tended to lower the rating. A large number of *dd* responses was a much more negative sign than a large number of *dr*. Imbalance between the amount of animal movement and the use of texture (few *FM* responses and many *Fc* or many *FM* and few *Fc)* lowered the rating. Recoverability after "shock" reactions moved the rating toward the "adjusted" end of the scale. The use of a variety of content was considered a positive sign. The judges agreed that the primary use they made of the content was to indicate basic conflicts and current upsets.

Particular aspects of the Rorschachs were emphasized differentially by the three judges. Judge C was more influenced by the content than the other two raters. She was concerned with the freedom or constriction of verbalizations and seemed to use the test more like an interview than the other judges. She gave more emphasis than A or B to *M's* as a positive sign and to stereotypy and lack of color as negative signs. Thus, for Judge C, an "adjusted" Rorschach meant an "ideal" or mature record rather than a test which was average for this group.

Judge B conceived of adjustment in this group as being characterized by the popular or usual response. The absence of certain variables or responses common to this group lowered the adjustment rating. The use of the Klopfer-defined populars was especially considered, particularly the handling of the animals on Card VIII as representative of self-acceptance. Along with Judge A, Judge B felt that immaturity does not mean "maladjustment" for this group. Nor was a temporary upsetness, such as might be shown in Card VI, enough to give a "maladjusted" rating at this age.

Even more than Judge B, Judge A built her adjustment scale around the characteristics of this group. Thus, absence of color responses, stereotypy, and constriction of verbalization did not in themselves lower the adjustment ratings. However, inconsistencies in these areas, such as the absence of color-determined responses in combination with a high VIII-X%, did indicate conflict for her and consequently lowered the adjustment rating. Judge A was especially conscious of assigning differential weightings to many Rorschach variables and relationships in male and female records. For example, while *K* and *FK* were almost invariably negative signs in a boy's record, they were not necessarily negative in a girl's. Or a large number of texture-determined responses in combination with few animal-movement responses lowered a boy's rating more than a girl's.

It is interesting to note these varying approaches of the judges even though this study does not purport to present an extensive analysis of the behavior of Rorschach interpreters. Differences among the "experts" should not be overemphasized, however. It was as evident from the conference of judges as from the correlations and per cent agreements presented above that the raters approached their task with much the same Rorschach attitudes, and that the differences occurred primarily in the hierarchy of cues rather than in the cues themselves.

The notes made by Judge D, the "naive" rater, mentioned the following qualities as being important in determining her ratings: imagination, quality of observation, originality, ease of verbalization, affect represented, capacity to use and modify what Judge D interpreted as "normal" responses, defensiveness about the test and the concepts seen, and degree of satisfaction or dissatisfaction with responses.

Some of the spontaneous recorded comments of Judge D would be of interest to Rorschach-trained people, as they show a startling correspondence to classic Rorschach terminology and interpretive practice. For example, of a record with many good *M's*, she notes, "Enjoys her acquired knowledge. Rich inner life; reactions of self of importance to self." Of another with many white space responses, she says, "Goes out of way to see things, ignoring the close-at-hand while doing so. Sounds like belligerence; looking for trouble."

Judge E, the "novice" rater, used content a great deal, "as in a TAT." The affect expressed by the subject was important in making her ratings, as were the kinds of *M's* and color responses. Responses given to infrequently used location areas lowered her ratings.

6. *Relationship of judges' ratings to Davidson Adjustment Signs* (1)

In the process of making explicit their cues, the "experts" often justified the adjustment ratings they had given by listing many of the Rorschach relationships included in the adjustment "sign" lists. This was true even though the judges were aware of nonsignificant relationships between the "signs" and interview ratings of adjustment (5). They

TABLE 6

CORRELATIONS BETWEEN JUDGES' RATINGS AND THE DAVIDSON ADJUSTMENT SIGN LIST*

Judge	Boys ($N=71$)			Judge	Girls ($N=75$)		
	Sort				Sort		
	1	2	3		1	2	3
A	.37	.39	.37	A	.45	.29	.43
B	.52	.53	.52	B	.40	.56	.37
C	.28	.31	.39	C	.36	.23	.35
D	—	—	—	D	(C girls, $N=38$)		.12
E	(G boys, $N=41$)		.59	E	—	—	—

* Significance of these biserial correlations is indicated as follows:
Significantly different from zero at 1 per cent level, ═════
Significantly different from zero at 5 per cent level, ─────

did not seem to use the "signs" in an additive fashion, in which each additional "sign" present raised the adjustment rating. The context in which the "sign" occurred determined its importance as an adjustment indicator. Yet, since the "signs" were so often mentioned in the judges' conference, we were curious to know how the ratings correlated with the "sign" scoring on the same records. The correlations are presented in Table 6.

We have evidence that the "signs" are being used to some extent, since the r's are all positive. But many other cues also seem to play a part in determining the final subjective rating. Because the Davidson list is based primarily on the formal aspects of the test, we would expect to find the highest correlations with the judges' first sort, when Rorschach content was not being used. The table shows no consistent change over the three sorts, however. The impression of Judge C, that she was least influenced by the formal aspects of the test and most influenced by the content, is confirmed here. Her correlations with the "signs" are slightly lower than A's or B's.

As would be anticipated, the Rorschach ratings of Judge D correlate lowest of all with the formal "sign" list. The correlation for Judge E is in line with those of the "experts."

B. VALIDATION

1. *The criteria*

Throughout this section, the G and C groups will be discussed separately, since the outside criteria of adjustment consist of ratings made on a somewhat different basis for the two groups. The primary differences concern the intensiveness with which the cases in the two groups were studied. Both G and C groups took part in the same testing program. In addition, the G groups were much more intensively interviewed and received yearly ratings on many personality dimensions, including the scales of adjustment. For purposes of correlation with our judges' ratings from Rorschachs, two case workers who had known and interviewed the G group subjects for many years were asked to make current adjustment ratings on those cases with which they were most familiar. Their ratings were *estimates of current adjustment* on four of the scales (self-adjustment, social adjustment A, social adjustment B, and family adjustment) which had been rated at each yearly interview. Thus, the G

group case workers had available as a basis for their ratings transcripts of yearly interviews with the subjects and previous ratings on the same scales.

The interviews on which the ratings were made might be called "diagnostic," in the sense that certain basic areas were systematically sampled each year and then evaluated in codified form so that standard material was obtained on every child. Thus, the ratings were based both on behavioral information reported by the mother, the child, and the teacher, and on the observations and inferences of the case worker regarding relationships and attitudes. The coding was governed by an over-all description of the area to be coded and by an individual definition of each of the points on the five-point rating scale. These code book descriptions and definitions are presented in Appendix A.

For the *C* group, special current ratings were made by the case worker who had known the subjects over the entire span of the study. These ratings were made using the same definitions and codes of adjustment, but were based on more casual observation of the subjects during testing sessions at the Institute and objective reports concerning the functioning of these individuals in the home, in the school, and on the job.

The distribution of the case workers' ratings for the four adjustment scales is given in Table 7. In contrast to the Rorschach adjustment ratings, they show a fairly normal distribution of the ratings over the five scale points. The ratings of the *C* boys (previously mentioned by the Rorschach judges as the group having the most disturbed content) show some skewness with a high percentage being rated 1 or 2 on all of the scales except family adjustment. Of the *C* girls, 56 per cent are rated as having very good (ratings of 4 or 5) family adjustment.

Table 8 gives the intercorrelations among the case workers' ratings on the four different scales of adjustment. All of these correlations are significantly different from zero in a positive direction. There are probably two factors involved: some halo which is operating in the case worker's evaluation of each child, and a certain degree of true relationship among the various areas of adjustment. These high intercorrelations among the four ratings on each subject

TABLE 7

PERCENTAGE DISTRIBUTION OF CASE WORKERS' RATINGS OF ADJUSTMENT

Rating	Self	Soc. A	Soc. B	Family	Rating	Self	Soc. A	Soc. B	Family
	G Boys ($N=41$)					*G* Girls ($N=37$)			
5	2%	10%	2%	7%	5	3%	8%	3%	3%
4	32%	32%	27%	24%	4	24%	22%	24%	19%
3	22%	34%	49%	41%	3	30%	43%	41%	43%
2	41%	24%	24%	20%	2	35%	19%	27%	27%
1	2%	0%	0%	7%	1	8%	8%	5%	8%
	C Boys ($N=30$)					*C* Girls ($N=38$)			
5	13%	3%	10%	27%	5	11%	5%	5%	24%
4	20%	20%	17%	7%	4	34%	5%	13%	32%
3	20%	40%	33%	40%	3	24%	61%	42%	24%
2	33%	37%	40%	20%	2	21%	24%	37%	16%
1	13%	0%	0%	7%	1	1[illegible]%	5%	3%	5%

TABLE 8

INTERCORRELATIONS AMONG THE CASE WORKERS' RATINGS ON FOUR ADJUSTMENT SCALES*

	G Boys					*G* Girls			
	Self	Soc. A	Soc. B	Family		Self	Soc. A	Soc. B	Family
Self		$\underline{\underline{.60}}$	$\underline{\underline{.61}}$	$\underline{\underline{.58}}$	Self		$\underline{\underline{.62}}$	$\underline{\underline{.64}}$	$\underline{\underline{.42}}$
Soc. A			$\underline{\underline{.73}}$	$\underline{\underline{.45}}$	Soc. A			$\underline{\underline{.70}}$	$\underline{\underline{.35}}$
Soc. B				$\underline{\underline{.53}}$	Soc. B				$\underline{\underline{.49}}$
	C Boys					*C* Girls			
Self		$\underline{\underline{.59}}$	$\underline{\underline{.75}}$	$\underline{\underline{.75}}$	Self		$\underline{\underline{.71}}$	$\underline{\underline{.74}}$	$\underline{\underline{.59}}$
Soc. A			$\underline{\underline{.77}}$	$\underline{\underline{.55}}$	Soc. A			$\underline{\underline{.86}}$	$\underline{\underline{.68}}$
Soc. B				$\underline{\underline{.64}}$	Soc. B				$\underline{\underline{.66}}$

* Significance of these Pearsonian correlations is indicated as follows:
Significant at the .01 level of confidence, $\underline{\underline{\qquad\qquad}}$
Significant at the .05 level of confidence, $\underline{\qquad\qquad}$

will lead to few differences in correlation of the Rorschach ratings with the different areas of adjustment rated from interviews.

2. *The results*

a. Gross correlations with the criteria. Figure 1 gives the Pearson correlations between the Rorschach ratings of adjustment and the case workers' ratings of adjustment for the four groups of subjects. A glance at this figure suffices to establish the most general interpretation of our results—namely, that there is no evidence of a significant positive relationship between the Rorschach ratings and the other measures of adjustment. Of the 144 correlations represented in the figure, only six are significantly different from zero at the 5 per cent level of confidence. The majority of the correlations cluster around zero, two-thirds of them in a positive direction.[4]

This low positive trend in correlations between interview-judgments and judgments based on projective techniques is a frequent finding in clinical research. It is in line with the tentative results of the research in assessing Veterans Administration trainees now being conducted at the University of Michigan (6). Guilford (3, p. 5), in discussing the extensive research in the aviation service during the war, says:

> One general conclusion that was brought home to us by repeated experience is that we should have greater respect for low correlations. Tradition has taught us that unless coefficients of correlation are substantial, for example .40 or above, there is too little relationship to bother with. We must face the fact, unpleasant though it may be, that in human behavior, complex as it is, low intercorrelations of utilizable variables is the general rule and not the exception.

---

[4] Between the ratings of the Rorschach "novices" and the criteria, a similar pattern of correlations prevails. Thus, the ratings by Judge E of the G boys correlate —.01 with the case workers' ratings of self-adjustment, +.33 with social adjustment A, +.33 with social adjustment B, and —.10 with family adjustment. The two +.33 correlations are significant at the .05 level. The ratings by Judge D of the *C* girls correlate +.11 with the caseworkers' ratings of self-adjustment, —.03 with social adjustment A, —.05 with social adjustment B, and +.03 with family adjustment. (These correlations, based on ratings by the Rorschach "novices," are not included in Figure 1.)

TABLE 9

Results of Analysis of Variance of the z's from Each of the Four Groups

Source	Σ of Squares	*df*	Variance	*F* Based on Triple Interaction Variance	Level of Confidence	*F* Based on Population Variance	Level of Confidence
			G Boys ($N=41$)				
Sorts	.041208	2	.020604	6.45	.01		
Interview Scales	.706404	3	.235468	73.75	.001	8.95	.001
Judges×Sorts	.318733	4	.079683	25.44	.001	3.08	.05
Judges×Interview Scales	.131260	6	.021876	7.86	.001		
Triple Interaction Variance = .0040426					Population Variance = .026316		
			G Girls ($N=37$)				
Sorts	.095838	2	.047919	5.32	.05		
Interview Scales	.224327	3	.074775	8.28	.001		
Triple Interaction Variance = .0098249					Population Variance = .029412		
			C Boys ($N=30$)				
Judges	.039210	2	.019605	4.51	.05		
Sorts	.091696	2	.045848	10.51	.001		
Interview Scales	.214182	3	.071394	16.41	.001		
Judges×Sorts	.209501	4	.052375	11.33	.001		
Triple Interaction Variance = .0041409					Population Variance = .037037		
			C Girls ($N=38$)				
Judges	.243016	2	.121508	56.11	.001	4.28	.05
Sorts	.071579	2	.035789	16.52	.001		
Judges×Sorts	.208117	4	.052029	23.90	.001		
Triple Interaction Variance = .0028245					Population Variance = .028571		

pect much difference among the judges in the matter of validity.

(2) *Differences among the three sorts in correlations with the criteria.* Variance from one sort to another is significant in all four groups. Obviously, however, it is not in a consistent direction. The correlation with the criteria may drop with one judge from the first to the third sort and with another increase, while these same judges' correlations will reverse their direction in another group. Increasing amounts of Rorschach information, then, did not result in increasing (or decreasing) correlation with the criteria.

The only interaction which is significant for three of the four groups of subjects is that of judges and sorts. In other words, the judges are acting differently on the different sorts in terms of their correlation with the criteria. That this is the case is obvious from inspection of Figure 1.

(3) *Differences among the interview scales in correlations with the Rorschach ratings.* There is one *F* significant at the .001 level of confidence when we use the population variance as the divisor. This is variance among the correlations with the four different interview scales for the G boys. The majority of the G boys were rated by one case worker who apparently (see Table 8) made greater discriminations among the four types of adjustment than the other raters. Using the triple

interaction variance in the *F* ratio, the correlations vary significantly among the four interview scales for all groups except the *C* girls. An inspection of Figure 1 will show that the judges' Rorschach ratings correlate more highly with the case workers' ratings on the scales of social adjustment than with ratings on the other adjustment scales. All of the significant correlations are with the two social adjustment scales, while the majority of the correlations with family adjustment are low negative.

Combining the impressions from all of the data, one is led to the conclusion that perhaps the most interesting aspect of the study is not the subjects, their Rorschachs, and their case histories, but the workers themselves, the Rorschach judges, "expert" and "naive," and the case workers. Had we limited our study to *G* boys only, for example, our conclusions might have been clearer and more confident than they are when we check them with results from the other three groups also. One major difference among the groups arises from the case workers' ratings of the subjects. For example, the mean interscale correlation for the worker who rated most of the *G* boys was .50, whereas the mean for the worker who rated all of the *C* cases was .70. This fact alone, since the Rorschach judges rated on only one scale, would influence the difference in correlation with the four scales which shows up in Figure 1.

c. *Case studies.* In a further attempt to explore the relationship between the two types of ratings, some individual cases with wide discrepancies were studied. There were seven cases whose Rorschach ratings were in the fourth quartile and whose interview ratings were in the first quartile. There were six cases whose Rorschach ratings were in the first quartile and interview ratings were in the fourth. The histories of those discrepant cases from the *G* group were thoroughly perused without any hypotheses in mind, in the hope that they might suggest hypotheses which could be checked against the cases from the *C* group. To a certain extent (even more than one might expect with such a small number of cases), this was successful.

Five of the seven cases whose Rorschach ratings were much higher than the case workers' ratings had very similar family patterns. These included desperately severe economic insecurity during childhood, much marital conflict between the parents with threatened separation and divorce, and siblings who, under these conditions, became delinquent. In all of these cases our subjects, although they certainly had problems and tensions, might be described as the best adjusted members of very poorly adjusted families. Perhaps their "good" Rorschachs reflect the capacity to make some sort of adjustment to an almost impossible environment. The case workers' ratings, on the other hand, used the whole study group as a reference population, i.e., the subjects were rated on their adjustment relative to the other subjects in the group, *not* relative to the job of adjusting they had to do. It is interesting too that absorbing interests in music or religion or both were common to all of these subjects and may have been important factors in their adjustment.

The pattern in the six cases whose interview ratings were high and Rorschach ratings low is not quite so clear. Three of these subjects have in common a high socioeconomic status, which has resulted in smoother life circumstances for them than for the children in the other group. There is no doubt that they are attractive young people, socially successful in a school system with definite social hierarchies. The conflicts resulting in a "poor" Rorschach have perhaps not been precipitated and resolved in this smooth, non–tension-producing environment. Another two cases are quiet and unanimated, with relatively low IQ's for our group and a low level of aspiration; they have accepted their personal and social liabilities and appear to the case worker on these grounds to be adjusted. To the Rorschach worker, however, real adjustment is questioned because their Rorschachs seem flat and withdrawn.

V. SUMMARY

The Rorschach records of 146 normal 18-year-olds (71 boys and 75 girls) were rated for over-all adjustment on a four-point scale by three Rorschach "experts." Each of the records was rated three times by each of the raters. In the first rating or "sort" of the subjects, only the formal scoring categories of the Rorschach were used; in the second rating or sort, the distribution of these categories over the ten cards was included; in the final rating or third sort, the whole protocol was used. Data from the three rating trials or "sorts" were separately treated in this study.

The ratings were made with the subjects divided into four groups: 41 of the boys (*G* boys) and 37 of the girls (*G* girls) had been intensively studied throughout their lifetimes as members of the Guidance Study of the Institute of Child Welfare, University of California. The remaining 30 boys (*C* boys) and 38 girls (*C* girls) were members of a matched control group in the Guidance Study and had been studied less intensively over the same period of time. For purposes of validating the Rorschach ratings against ratings of adjustment made by case workers, these four groups were treated separately, since the amount of information available to the case worker as a basis for adjustment ratings differed in the *G* and *C* groups.

The Rorschach judges consistently placed from 61 to 71 per cent of the cases on the "maladjusted" end of the four-point scale. However, the magnitude of the inter-rater agreement correlations showed that "experts" can agree, not only among themselves but also with judges having little or no Rorschach training. (Two other judges, one of whom had only a few months of Rorschach training and one of whom had absolutely none, rated a small group of records.) In verbalizing their cues, the "experts" mentioned many of the formal Davidson adjustment signs (their ratings correlated +.23 to +.56 with the Davidson list), but indicated that the signs were weighted with regard to the total context and that many other less objective cues were also used.

Concerning inter-rater agreement, there was no particular advantage or disadvantage to using the full Rorschach protocol over the formal aspects alone. With the addition of content, the raters changed more of the girls' ratings than the boys'. There was more consistent inter-rater agreement on boys' records than on girls' within a single sort, and more consistent intra-rater agreement on boys' records than on girls' between sorts. More boys than girls were rated as "maladjusted." This is in line with the results of our analysis of "neurotic" signs (5), in which the boys consistently received higher scores, and the analysis of adjustment signs (5), in which the boys received lower scores.

Correlations of the Rorschach ratings with the case workers' adjustment ratings were primarily low positive. The lack of a statistically significant relationship can lead only to the conclusion that the Rorschach, used blindly (i.e., without supplementary background and observational data) is of little value in predicting adjustment ratings made primarily on the basis of interviews and case histories.

Within this framework of no validating relationship, however, a few interesting trends emerge. Those Rorschach ratings which did correlate significantly with the criteria were with ratings of social adjustment. The correlations with family adjustment tended to be negative. Discrepancies among the results with the different workers point up the difficulty of using clinical judgments as validating criteria. A few individual cases with wide discrepancies between the two types of ratings were examined and interesting personal and sociological concomitancies observed. These concomitancies are worthy of verification by further investigation. As yet, however, the sources of the discrepancies between the Rorschach ratings of adjustment and the interview ratings remain undetermined. It is suggested that differing social background of the subjects may be responsible for some of the discrepancies; sociological biases of the case workers may also be influential; and finally, of course, there is the possibility of inadequacies both in the Rorschach and the interview data, as such.

VI. CONCLUSIONS

After hearing of the results of this study, an academic psychologist said to the authors, "*Now* are you finally ready to give up the Rorschach?" That, of course, is too simple. We know that a Rorschach interpretation can add meaning to the understanding of an individual case. But it is true that our confidence in the ability of Rorschach workers to analyze records of normal subjects for use in group research has suffered.

The authors would like to suggest five possible sources of the negative results in our validation attempts, only one of which is the Rorschach itself.

1. One of the difficulties may lie in the nature of our sample. It is primarily a normal sample with no pathological extremes, and finer discriminations than are obtainable by a brief four-point scale are necessary to differentiate homogeneous samples.

2. The Rorschach may represent such a limited behavior sample, so sensitive to temporary disturbances, that it is not useful for predicting "adjustment" in other situations or at other times. Or the test may pick up such evidence of instability as exists in the individual without ever tapping his areas of stability. The Rorschach may offer us no picture of present adjustment but rather predictions which only time and situational concomitancies can test. This projective technique may be one which will give us knowledge only of the subject's inner potentialities and functioning, never of his overt behavior. Another point which arises involves the theoretical question of the function of fantasy. If one subscribes to the theory that fantasy offers a "safety valve" device for many individuals, then, of course, we would expect no one-to-one relationship between the pictures of adjustment presented in the overt and covert areas.

3. The third possible source of our negative results is the Rorschach rater. A large part of Rorschach training consists of learning to recognize signs of abnormality which have been established

on pathological samples, with little attempted comparison with randomly selected normal groups. Our raters have in common with other Rorschach workers this background of training and vicarious experience through the literature, which has sensitized them to the conventional indicators of maladjustment. Evidence of this concern is to be found in the distribution of their ratings in which the majority of cases were rated "maladjusted." However, the positive correlations of our "naive" rater with the other judges and her zero correlations with the validating criteria indicate that the problem does not lie completely in stereotyped thinking and training on the part of our "experts."

4. A fourth series of problems concerns the criterion variables used, namely, case workers' ratings of adjustment. Although all case workers used the same rating scales, they were often of different theoretical orientation. While all had access to previous case records, they varied with regard to the length of time they had known the subjects, the degree of openness of the relationship, and thus with regard to their understanding of the dynamics of the cases. The difficulty posed by differences between the *G* and *C* groups with regard to intensiveness of knowledge about the cases further complicates our attempts to pin down the factors involved in our negative results.

5. A fifth possible factor in the interpretation of our results is the semantic difficulty inherent in the use of the concept of adjustment. The concept of overall adjustment, used by the Rorschach raters, is too broad and too vaguely defined. The meaningfulness of so inclusive a concept of adjustment, without regard to time, place, or area of functioning, may be questioned. Further, the concept of adjustment used by the case workers was specified as to these variables, while that used by the Rorschach raters was not.

These considerations have some implications for future validation attempts with the Rorschach. Research clinicians using this tool might well spend more time and effort collecting and studying the records of adequately functioning individuals, in order to counteract their biases in the direction of abnormality. The results also suggest that narrower, much more specific areas of personality be studied through the Rorschach for ease in finding suitable validating criteria. Another point, which has been made many times before, might be reiterated here. Much more meaningful results may be obtained when we set up our experiments so that the Rorschach is not tested in isolation, but rather investigated in terms of the additional predictive significance which it gives in combination with other clinical measures.

APPENDIX A

Adjustment to self (size, sex, physical make-up, abilities, disposition, past misdemeanors, etc.)

(1). Extreme lack of confidence shown either through excessive self-depreciation, unfavorable comparison, or marked braggadocio. Intense guilt feelings, or feelings of inadequacy.

(2). Not as marked as 1 but marked unacceptance of one or two items of make-up. However, if to the point of dominating whole personality, classify under 1. Or many mild points of discomfort. Over-adjusts to others at own expense.

(3). Lacks confidence in some particulars but for the most part is unself-conscious or unaware of assets or liabilities. Occasional tension or conflict about self but not characteristic.

(4). Recognizes handicaps but accepts them realistically and without much tension. Little evidence of compensatory behavior.

(5). Full of confidence—accepts limitations and assets easily.

Adjustment to family (codings on the severity of tension—whether it involves one or more family members, including step-parents. Tension-producing relationships whether shown openly or by withdrawing or in compensatory demonstrativeness or such intense attachments or such ingrown relationships that normal, friendly relations to outsiders are precluded).

(1). Marked hostility or insecurity with parents, feels he is not loved, discriminated against, or fails to meet parents' expectations or sibs' accomplishments. Humiliation over parents' faults, occupation, income, etc. Child subjected to rather intense, ambivalent swings. Practically no compensatory satisfaction.

(2). Same trends, but not so extreme as in 1. Much more strain than average but some sources of security. Very poor relationship with one parent. If this poor relationship is with a parent with whom he seldom comes in contact, it should be included here *unless* it dominates his other family feelings, in which case the coding is 1.

(3). For the most part, low-grade tension; not markedly unsatisfactory, but enough insecurity in relationship to family members to make for strain or chronic low-grade embarrassment over family members; mild jealousies.

(4). Absence of real strain—occasional episodes of insecurity or lack of outstanding satisfactions.

(5). Complete security in family affection, parents, sibs. Genuinely proud of parents or a realistic (unembarrassed) acceptance of parents' faults, limitations, or assets. No evidence of tension-producing attachment.

Social adjustment

A. (Refers, as used here, to social techniques, ease with which a child evokes acceptance, approval, friendliness or loyalty; or incurs disapproval, distrust, or hostility. Emphasis is on a social criterion rather than on child's security in social situations—this last being handled under such items as shyness, friendliness, quarrelsomeness, etc.)

(1). Extremely unsuccessful in social relations. Intensely disliked or resented by practically all of his playmates. Very poor social techniques and constantly does things which offend his social group. Actively avoided.

(2). Unsuccessful in social relationships; not popular but does not give rise to the intense reactions of 1. Ignored or not spontaneously included in group play, etc. Does things which annoy. Definitely disliked by many children; has some supporters.

(3). Not outstanding. Normally successful in having people like him. Some playmates may not like him but has approving supporters. Or successful with one sex but unsuccessful with the other.

(4). More popular than the average; has good social techniques.

(5). Unusually successful in social relations. Inspires friendliness or loyalty. Has many people devoted to him. Constantly gives rise to such statements as, "Isn't he a swell guy?" "She's a grand person," etc.

B. (Refers, as used here, to child's feeling of tension or confidence and satisfaction in social relations)

(1). Acutely uncomfortable, dislikes people, feels he is treated unfairly or is disliked by others. Chronically sulks or is aloof.

(2). Definite but less acute tension than 1. Feels he is not liked as well as other children. "Children don't play with me much." Or worried because one or two children dislike him. Insecurity may take form of shyness or aggressive attempts at social overtures.

(3). Takes his social relationships for granted, doesn't occur to him to consider whether he is liked or disliked. Some frictions but "have as many friends as most children do." May have few friends but apparently has no tension or worry on this score. Feels secure with one sex but a bit uneasy with the other.

(4). More confidence in social relationships than average. Very little tension and many satisfactions.

(5). Likes children very much; enjoys being with them. Is enthusiastic and happy in relations. Confident, feels socially secure.

BIBLIOGRAPHY

1. Davidson, Helen H. *Personality and economic background.* New York: King's Crown Press, 1943.
2. Dunlap, J. W., and McNamara, W. J. A graphical method for computing the standard error of biserial r. *J. exp. Educ.,* 1934, **2**, 274-277.
3. Guilford, J. P. Some lessons from aviation psychology. *Amer. Psychologist,* 1948, **3**, 3-12.
4. Hayes, S. P., Jr. Tables of the standard error of tetrachoric correlation coefficient. *Psychometrika,* 1943, **8**, 193-203.
5. Ives, Virginia, Grant, Marguerite Q., and Ranzoni, Jane H. The "neurotic" Rorschachs of normal adolescents. *J. genet. Psychol.,* 1952.
6. Kelly, E. L., and Fiske, D. W. The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist,* 1950, **5**, 395-406.
7. Klopfer, B., and Kelley, D. M. *The Rorschach technique.* Yonkers, N.Y.: World Book Co., 1942.
8. Macfarlane, Jean W. Studies in child guidance. I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.,* 1938, **3**, No. 6.
9. McFate, Marguerite Q., and Orr, Frances G. Through adolescence with the Rorschach. *Rorschach Res. Exch.,* 1949, **13**, 302-319.

VOLUME 66
NUMBER 3

WHOLE NO. 335
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Some Personality Correlates of Religious Attitudes, as Determined by Projective Techniques

By

RALPH MASON DREGER
Florida State University
Tallahassee, Florida

Accepted for publication February 12, 1951

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

ho

E

COPYRIGHT, 1952 BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

TABLE OF CONTENTS

Preface iv

I. *Purpose* 1

II. *Summary of Research* 1

III. *Selection and Use of Tests* 2

IV. *Experimental Procedure* 3

V. *Quantitative Methods and Procedures* 8

VI. *Conclusions* 14

Quantitative comparisons 14

Qualitative comparisons 16

VII. *Summary* 17

Bibliography 18

PREFACE

This study was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Southern California, June, 1950.

Acknowledgements are made to the writer's doctoral committee, Professors David D. Eitzen, Chairman, J. P. Guilford, Paul Irwin, Eric Titus, and S. Medford Wesley, who guided the dissertation. Thanks are also expressed here to the sixty subjects comprising the human raw material of the research. Without the willing cooperation of these persons the study would have been impossible.

SOME PERSONALITY CORRELATES OF RELIGIOUS ATTITUDES, AS DETERMINED BY PROJECTIVE TECHNIQUES

I. PURPOSE

THE ORIGINAL purpose of this study was to determine the "emotional maturity" of religious conservatives and liberals by means of projective techniques. Willoughby (21), Woodward (23), Symington (16), and Shrout (14) have made investigations by means of questionnaires of the emotional characteristics of religious persons. With the exception of the last, these investigators have concluded that generally liberals are more "emotionally mature" than conservatives. According to Shrout's findings, however, the degree of religious attitude toward God and the nature of a person's religious experience (in our society and culture) has little or no bearing on his score on the Willoughby "Emotional Maturity" scale.

Since the writer found the concept of "emotional maturity" loosely defined, the aim was adopted of comparing religious liberals and conservatives on specific personality variables which authorities consider to be measures of adjustment or maturity, emotional, intellectual, or social. An attempt was made to sample widely from such characteristics, presumably measured by the tests which were employed in the study.

As might be expected, the terms "conservative" and "liberal" are likewise inadequately defined for scientific usage. Therefore, two groups of persons were selected for the proposed comparison, differing not only significantly but widely from each other on two religious attitude scales; so that "conservative" and "liberal" are understood here to mean "as defined by the two selective tests" (described in Section III).

Hypotheses which were set up for guidance of the study were stated as follows:

1. *Religious liberals are more "emotionally mature" than religious conservatives.* "Emotionally mature" was taken as "maturity in general." It was considered that a favorable balance on the liberal or conservative side of a number of personality factors accepted by most authorities as representing adjustment would tend to prove or disprove this hypothesis. These factors were specifically stated in terms of scores on the personality tests employed.

2. *Rigidity of personality structure and conservative religious attitudes are positively correlated.*

3. *Different types of emotional control are found in both groups,* so that the same external type of behavior, classed as mature or immature by both subject and observer, will be found to issue from different forms of emotional organization.

4. *Conservative individuals are more guilt-ridden than liberals.*

5. *Liberals tend to gloss over aggression-provoking situations more than conservatives.*

6. *A greater need for dependence is found in the conservative.*

II. SUMMARY OF RESEARCH

An outline of the over-all procedure employed to test the above hypotheses is presented here:

1. A preliminary selection of churches representing conservative, liberal, and "mixed" congregations, total of 21.

2. Securing of cooperation of pastors and/or leaders for nonselective lists of

persons in churches.

3. Distribution of Salvation Opinionnaire by mail and directly through church groups, total of 1035.

4. Simultaneous distribution of blank for identifying information.

5. Division of the first group of respondents, 351 persons, into two main "populations" representing liberals and conservatives. Selection of possible subjects from persons about the submeans.

6. Administration of Ferguson (religious attitude) scale and Wonderlic Personnel Test (intelligence) to possible subjects to obtain 30 subjects for each group. (Elimination of 20 persons out of 80 possible subjects on the basis of low scores on Wonderlic, or on the basis of Ferguson scores not in accord with Salvation Opinionnaire; these persons called "eliminees" hereafter.)

7. Administration of Rosenzweig Picture-Frustration Study, Thematic Apperception Test, and Rorschach to subjects.

8. Transcribing, scoring, etc., of test protocols; statistical treatment of scorable data.

The location of the research was in Southern California in two large, practically contiguous cities, representing populations of considerable variety of background. Research was conducted in the summer of 1949.

Religious populations were found in 21 churches, 17 in one city and 4 in the other. Denominations included are: Baptist, Church of God, Congregational, Friends (or Quakers), Lutheran (Missouri Synod and American Lutheran), Methodist, Mission Covenant, Presbyterian, Unitarian, and Universalist. Independent churches include: Pentecostal Tabernacle-type (Baptist doctrine), and Unity-type (Theosophical–semi-Christian Science). All but the first and the last above contributed at least one subject to the study.

Identifying information concerning the 60 persons who participated as subjects of research is given in Tables 1 and 2. Note that in spite of an attempt to keep subjects comparable, Liberals' mean age and educational level differed significantly from those of Conservatives at the 5 per cent level. Conservatives' mean occupational rating (higher mean signifies lower status) was significantly higher than Liberals' (2 per cent level). It is also to be noted that men and women were distributed evenly in each group.

III. SELECTION AND USE OF TESTS

Since the study was to depend exclusively on tests, the selection of tests was considered important. The Salvation Opinionnaire has been satisfactorily validated both by a chi-square item analysis on a large population and by the judgment of qualified individuals (19). Ferguson's Primary Social Attitude No. 1, developed by means of factor analysis (7, 8) and administered to two independent student populations, was found to denote what Ferguson believes is an operationally stable factor, Religionism. The Pearson product-moment r between these two measures on the writer's subjects and eliminees together is .73; on subjects alone r is .77. Both of these coefficients are significant at the 1 per cent level.

Form F of the Wonderlic Personnel Test was found to be satisfactory from several standpoints: its ease of administration; its twelve-minute time limit; its adequate reliability (r: .88 to .94 on odd-even items on different forms); its validity in both practical personnel selection and compared with its parent, the Otis Self-Administering Test of Mental Ability, Higher Examination ([illegible] .81 to .87); and lastly, its disarming title (22).

To obtain as wide a sampling of personality structure and functioning for measurement of "maturity" factors, the Rosenzweig P-F Study, the Rorschach, and TAT cards 2, 3BM, 3GF, 4, 6GF, 8BM, 11, 14, 15, and 16 were selected.

Exclusive utilization of tests makes the conclusions of the study dependent on the validity and reliability of the instruments as well as the reliability of the scorer. None of the projective tests employed has proved itself to the same degree of statistical satisfaction that other more "objective" tests have proved themselves, although the clinician believes they are sufficiently valid and reliable to be useful.

Bernard (3) has conducted some reliability studies on the Rosenzweig. Tomkins (17) and Bell (2) present references to many studies which can be considered validation attempts for the TAT. Bell (2) summarizes validity research on the Rorschach.

As to the use of these tests, the Salvation Opinionnaire was sent out through the mail or administered directly to Sunday-school classes or other church groups. On the first contact with potential subjects the writer administered the Ferguson, Wonderlic, and Rosenzweig. In two instances the Rorschach and TAT were likewise given on first contact, but in most cases they were administered in two sessions. Administration was standard according to Rosenzweig's, Klopfer's, and Murray's instructions, with the exception that only on the first two cards of the TAT were the subjects reminded of the time. Protocols were manually recorded with a relatively efficient mongrel shorthand the writer has developed. Tests were given privately in any one of three offices or in the homes of subjects.

TABLE 1

IDENTIFYING INFORMATION ABOUT SUBJECTS

Item	Conservatives	Liberals
Number of subjects		
Male	15	15
Female	15	15
Marital status		
Married	27	23
Single	3	3
Divorced		2
Widowed		2
Students	3	4
Offspring	48	59

IV. EXPERIMENTAL PROCEDURE

A. CONTROL OF NONEXPERIMENTAL VARIABLES

In planning research the factors which were considered relevant to the study were set down, to be controlled if possible. These factors with the control measures exerted are as follows:

1. *Race.* Only "white" churches were selected.
2. *Background (sectional, geographical).* The two cities are cosmopolitan to a higher degree than most cities in the country.
3. *Religious tradition (Catholic, Jewish, Protestant).* Protestants only were selected. One person in each group is of Jewish background.
4. *Socioeconomic status.* Since incomes did not differ significantly (Table 2), even though occupational levels did to an extent, and social class differences were thus minimized (though, of course, other factors enter into social class distinctions), it seems reasonable to suppose that socioeconomic status would not significantly affect the experimental variables.
5. *Sex.* Fifteen persons in each group were male and fifteen female.
6. *Position in family.* The writer considered, perhaps wrongly, that this factor is not sufficiently relevant to require control, despite emphases of Adlerian dynamics.
7. *Age.* An attempt was made to keep subjects

TABLE 2

COMPARISON OF SUBJECTS ON IMPORTANT SELECTIVE FACTORS

Item	Mean	*SD*	*t*	Significance Level
Age				
Conservatives	33.87	7.90	2.02	5%
Liberals	38.63	9.94		
Education				
Conservatives	15.37	1.84	2.03	5%
Liberals	16.40	2.03		
Occupation[a]				
Conservatives	3.23	1.61	2.52	2%
Liberals	2.23	1.41		
Family Income[b]				
Conservatives	5.07	3.77	.89	—
Liberals	4.37	1.89		
Salvation Opinionnaire[c]				
Conservatives	7.84	.17	27.33	1%
Liberals	4.00	.74		
Ferguson[c]				
Conservatives	85.00	7.11	11.86	1%
Liberals	46.62	15.84		
Wonderlic[d]				
Conservatives	30.13	5.21	1.02	—
Liberals	31.77	6.89		

[a] Higher mean signifies lower status.
[b] In thousands of dollars.
[c] Religious attitude scales.
[d] Intelligence scale.

between twenty-five and sixty, a condition it was impossible to fulfill. The means differ significantly (Table 2), Liberals higher than Conservatives. Inspection reveals that the trend appears to be the same in the populations from which subjects were drawn.

To test whether age did have some influence on scores, a simple analysis of variance test was run to determine the influence of age and religious attitude on Rorschach *R* scores. Twelve cases were selected at random from each group for each of the age ranges, twenty to thirty-five and thirty-six to fifty. The results shown in Table 3 indicate that the influence of age itself is not important. The influence of attitude is very important, the *F* ratio being significant at the 1 per cent level. But also the interaction of age and attitude is significant at the 5 per cent level. One would suspect that both higher chronological age and liberal religious attitude together increase one's susceptibility to Rorschach stimuli.

8. *Intelligence.* Table 2 shows that the *t* ratio is not significant between the group means.

9. *Education.* Only high-school graduates or those with additional education were included as subjects. Even so, however, there is a significant difference between group means in education (Table 2). An analysis of variance was made to determine the comparative influence of age and education on total Rorschach responses (*R*),

TABLE 3

ANALYSIS OF VARIANCE TO DETERMINE INFLUENCE OF AGE, EDUCATION, AND ATTITUDE ON RORSCHACH NUMBER OF RESPONSES (R)

Factors Involved in Total Variance	Degrees of Freedom	*F*	Significance Level
Age and education	5 & 36	1.62	—
Age and attitude[a]	3 & 44	5.96	1%
Age only	1 & 44	2.18	—
Attitude only[a]	1 & 44	9.68	1%
Interaction of age and attitude[a]	1 & 44	6.05	5%

[a] Religious attitude: conservative or liberal.

dividing the entire sixty subjects into ages twenty to thirty-five and thirty-six and above, and using educational levels 12-14, 15-16, and 17-and-above. The *F* ratio is not significant (Table 3). One may possibly conclude that of itself the educational level did not significantly influence Rorschach *R*.

10. *Student status.* A minimum of students was thought best in order to have a representation of the adult population. Three students are found among Conservatives, four among Liberals.

11. *Marital status.* This variable would have been very difficult to control directly. Reference to Table 1 will show that the groups were somewhat comparable on marital status.

12. *Denominational background.* The original selection of religious populations and the limitation of numbers with respect to churches constituted an attempt to control this variable. In a larger sense denominational background means early training as well as present affiliation. It would have been an endless task to assure representativeness in this particular.

13. *Lay-clerical status.* Only lay persons were selected.

14. *Sunday school teacher-member.* It is assumed that a teacher might show different personality trends from a student's though this assumption is problematical. No attempt was made to select with this factor in mind.

15. *Religious attitudes.* This is one of the experimental variables. There is so much variation in definitions of "liberal" and "conservative" that the selection for this study was made operational to the tests mentioned in the first part of Section III.

16. *Church attendance.* The groups were not equated on this factor, even though it may be an important one from an experimental standpoint. Almost no churches keep an actual person-by-person record of church (worship) attendance. Church relatedness is all that can be postulated. Church membership was not considered as a variable, since in most churches it has a rather loose correlation with a person's actual relation to the church.

17. *Clinical normality-abnormality.* All persons were assumed to be clinically normal, actually functioning in society in a relatively normal way. Three persons in the Liberal group had had psychiatric (noninstitutional) treatment at some time.

18. *"Emotional maturity."* Maturity as a whole was one of the (multiple) variables to be measured.

19. *Observer's reliability as examiner and scorer.* This variable was held as constant as possible, with the researcher as the one who administered all tests except the Salvation Opinionnaire, scored both Rorschachs and TAT's, and made final decisions on the Rosenzweig.

20. *Validity and reliability of tests.* As indicated previously, these factors were assumed to be satisfactory.

21. *Form of contact with subjects.* One of the weakest links in the entire project is the manner of contact with prospective subjects. A delay forced by circumstances over which the writer had no control made contact with populations impossible at a time when Liberal churches were holding group meetings. The result was that twenty-five Conservatives were chosen from group distribution and five from letter distribution, while the numbers among Liberals were just reversed. A minute examination of the differences, if any, between group distributed and letter distributed may reveal some definite trends. But the research was continued as if both were the same, a possibly unjustified move.

22. *Volunteers.* Respondents to the Salvation Opinionnaire were asked to check if they would be willing to continue "in this very important religious project." Of the 490 who filled in the Opinionnaire 207 marked "Yes." In view of the selective nature of this factor, the study might be called, "An analysis of those who answer 'Yes' on religious questionnaires," or simply, ". . . of those who answer religious questionnaires."

Five of the Liberal subjects did not originally answer "Yes" on their returned letters. There were not enough Liberals who had checked assent who also came within the limits set on the Opinionnaire. So the writer had to pick at random three persons from the non-"Yes" group who did meet requirements; these persons manifested no hesitation in accepting a part in the research when the situation was explained to them. Two other non-"Yes" subjects were husbands of women who had consented to participate. No striking differences appear to be revealed between these five persons and the "Yes" group on personality measures, except that one of the five showed highly neurotic tendencies on the Rorschach and TAT, including four *C* on the former.

In the course of the experiment other variables came into prominence, chief of which were:

23. *Acquaintance with examiner.* Four of the Conservatives selected as subjects were known to the researcher from previous contacts; besides this he was known to some of the Liberals by reputation. It is doubtful that basic personality materials could be concealed or more fully revealed because of the acquaintance. Nevertheless, this factor may have exerted some uncontrolled influence.

24. *Husband-wife combinations.* Participating married couples may have affected each other's responses to some indeterminate degree. If so, the influence was probably balanced, as there

were four man-and-wife combinations in the Conservative group and three in the Liberal, with one couple split between Conservatives and Liberals.

B. SELECTION OF SUBJECTS

Altogether 1035 Opinionnaires were distributed; 490 were returned, of which twenty-two were defective. By a process somewhat similar to sequential analysis, representative subjects were selected prior to complete returns. The form of the first distribution from which subjects were selected did not differ significantly from that of the final frequency distribution when all returns were in.

Two major divisions were made at the mean of the entire original distribution, 6.788. Each major division was considered to be a distribution in itself, overlapping the other to some extent, with the lower division mean at 4·951, and the upper division mean at 7·799. These submeans fall in the modal intervals for Conservatives (high end of scale) and Liberals (low end of scale).

The two subject groups were then selected from within plus and minus one probable error of the submeans. Discarding after a brief trial a complex (and compulsive) system of limitations of ranges of individual scores and numbers of questions, the writer adopted a common-sense procedure of having prospective subjects in each group in rather close agreement with one another in range and number of scores on their Opinionnaires. Where persons had checked a larger number than twelve (out of twenty-four) statements, they were excluded, as were those with only one statement checked.

Insofar then as the total "sample populations" are representative of religious liberals and conservatives—and the discussion above throws some doubt on their representativeness—the "subsample populations" out of which subjects were drawn are definitely representative of liberals and conservatives.

C. UTILIZATION OF PERSONALITY MEASURES

When protocols of the sixty subjects had been obtained by the method outlined previously, scoring was completed (so simple it sounds—so many weeks it took!), and scoring categories which had been selected previously from the various tests and which tend to measure some kind of "maturity," were subjected to analysis.

On Rosenzweig's test the Group Conformity Rating scores were compared, even though Bernard (3) found little value in his studies of normals in the *GCR;* for it was assumed that if any normal-abnormal differences obtained between the groups investigated, the *GCR* would tend to bring them out. Reaction-type and direction of aggression columns were compared also, as well as the special cases of ego-defensiveness, $\underline{E}$ and $\underline{I}$, which constitute "denials" in Bernard's terminology. Since not everything on the tests could be utilized, "trends" were not analyzed; but on examination these may show significant differences between Liberals and Conservatives.

Beck (1), Klopfer and Kelley (11), and Mons (12) were consulted for Rorschach interpretations. Klopfer's scoring system was employed with few exceptions. A new weighted score was added (which did not prove significant with normals, but might with nonnormals), *"Sum K,"* comparable to *Sum C,* to represent the "insight-anxiety" continuum, as *Sum C* represents the "controlled responsiveness-uncontrolled responsiveness" continuum. Mons's psychograph (12, p. 92) suggests that the same kind of relation holds among *FK*, *k*, and *K* that holds for *FC*, *CF*, and *C* (also among *M*, *FM*, and *m*), with scores closer to the center of the graph indicating more favorable adjustments. *Sum K*, therefore, equals *FK* plus 2*k* plus 3*K*, all divided by 2.

For *A*, *Ad*, and *Aobj*, distinctions are made which are not made in Klopfer and Kelley (who are not clear at this point). *Ad* was scored as "parts of living animals and doubtfully living (usually *F*)." *Aobj* was scored as "parts of nonliving animals."

The modern movement in refinement of content analysis promises to be more discerning of personality characteristics than some of the more formal analysis. Nevertheless, the writer kept to more widely accepted scores and ratios, except for *Sum K*, and to the more easily quantifiable aspects of the Rorschach.

By far the most labor went into processing TAT protocols. Tomkins (17), Murray (13), and Stein (15) were reviewed and found unsuitable for complete quantification of data. Therefore, an adaptation of the "Thematic Apperception Test Analysis Guide" (Wesley, 20) was employed, a system which allows for the scoring of smaller units than Murray's or Stein's systems (while at the same time not scoring every word as with Tomkins). The "Guide" depends in part upon Murray's definitions.

Guetzkow (9) takes up problems of unitizing and categorizing data on qualitative tests. He points out that much of the interpretation depends on what units the interpreter selects to score, and on the category sets he employs. As far as the present research is concerned, units were fairly well defined by the system, as phrases, sentences, or in some cases a couple of sentences, units which could be taken as separate needs, press, inner states, or defense mechanisms. Where such a unit could not be scored with only one factor, that is, it expressed in the same words two dynamic trends or situations, the scorer did not hesitate to use two scoring factors. No attempt was made to score for levels (Tomkins), though it seems there is no reason why quantitative assessment could not be made of depth or level.

Altogether 273 factors were scored, some only a few times, but scored nevertheless. Therefore, separate scores were combined into what the writer believed to be operational units. These were combinations on theoretical bases, not on statistical analysis of separate scores.

The first 10 cases in each group did not receive scores on some of the factors. The reason is a purely personal and expedient one. At first the writer was scoring only on the scores which he had decided were significant to the research. But the procedure proved too frustrating. For the last forty cases all the scoring factors in the "Guide" and a few others were employed. Less frustration and less wasted time resulted from using more rather than fewer factors.

Since the change from scoring some to scoring all the factors also made a difference when transforming raw category scores into percentage scores (which were utilized for comparisons), when significance tests were run, tests for both thirty cases apiece and the last twenty cases apiece were made. It was assumed, however, that since omissions occurred in ten cases in each group, any discrepancy caused by the change would tend to balance itself out. In all cases except one, that designated Column 9,[1] *n Dep*, *−n*

---

[1] Defined below.

Dep (I), the assumption was borne out; in this case, the only seemingly unequivocal difference between the two groups of thirty cases apiece disappeared when twenty cases were compared.

Out of 49 combined-factor categories to which individual scores were assigned only 19 were compared, since the others did not yield enough scores to make significance tests. For the definitions of the individual scores comprising a category, reference will have to be made to the "Guide" already mentioned (Wesley, 20). But the combined categories are roughly characterized by the statements below; the reader may be able to follow the meanings of scoring symbols to some extent by comparing them with summaries. The scores were originally set up in columns, hence *are referred to as "Columns" below.*

Column 1: References to the external world, physical and mental capabilities, and opportunities in general, beneficial to hero. *(p Env, p Good, p Ment, p Oppor, p Phy)*
Column 2: Bad health, bad fortune, lack of opportunities, danger, or trouble in general, directed toward hero. *(p Afflic, p Bad, p Dang, p Troub, —p Env, —p Oppor)*
Column 3: Hero's feelings and attitudes ones of security, certainty, adequacy, calm, lack of turmoil and anxiety. *(Adeq, —Anx, —Conf, —Confl, Cert, —Dist)*
Column 4: Hero's feelings and attitudes ones of uncertainty, insecurity, inadequacy, turmoil and anxiety. *(—Adeq, Anx, Conf, Confl, —Cert, Dist)*
Column 5: Hero's self-blame, or failure to blame self. *(n Agg (I), —n Agg (I))*
Column 6: Hero's attacks on others, verbal or physical, including deception and rejection. *(n Agg, n Agg (P), n Decep, n Rej)*
Column 7: Hero's resistance to authority, or show of independence rather than dependence. *(—n Compl, —n Dep, n Dep (I), n Resis, —n Subm)*
Column 8: Hero's desire for or display of affection or friendship. *(n Affec, p Suc-Affec, n Affil)*
Column 9: Hero's dependency, or failure to trust in or depend on self. *(n Dep, —n Dep (I))*
Column 10: Hero's desire for or success in gaining freedom. *(n Free, p Suc-Free)*
Column 11: Hero's sympathy with, protection of, or help directed toward another. *(n Emp, n Help)*
Column 12: Hero's failure to show sympathy, give protection to, or help another. *(—n Emp, —n Help)* Column 13: Defense mechanisms employed either by subject or hero, or hero's defense of self. *(Def Mech, n Just)*
Column 14: Others' acceptance of, dependence on, affection for, or failure to reject or aggress against hero. *(—p Agg, —p Agg (P), p Affec, —p Decep, p Dep, p Affil, —p Rej)*
Column 15: Others' aggression against, rejection of, lack of dependence on, or failure to show *affection for hero. (p Agg, p Agg (P),—p Affec, p Decep, —p Dep, —p Affil, p Rej)*
Column 16: Others' restraint of or dominance over hero. *(p Coerc, p Dom, p Pers, p Proh, p Restr)*
Column 17: Others' sympathy with, protection of, or help directed toward hero. *(p Emp, p Help)*
Column 18: Failure of others to show sympathy with, give protection to, or help hero. *(—p Emp, —p Help)*
Column 19: Hero's striving for or success in achieving higher goals, money, education, or knowledge. *(n Ach, n Acq, p Suc-Ach, p Suc-Acq, p Suc-Cog)*

A number of individual scores on each test were specifically hypothesized in relation to various levels of maturity. Others were not. Only those which were hypothesized beforehand were allowed to influence conclusions directly. (See Section VI.)

V. QUANTITATIVE METHODS AND PROCEDURES

In reducing qualitative data to quantities some of the original quality is lost. Nevertheless, when the purpose is comparison of groups, quantitative comparisons are really the only accurate kind that can be made, for qualitative comparisons are too much subject to individual biases, too much influenced by the very individuals who should be partially disregarded in comparing groups, namely the individuals in the tails of the distributions.

Cronbach (4, 5) summarizes difficulties in making tests of significance on multiscore tests. For the Rorschach specifically he recommends raising the *P* value re-

TABLE 4

RORSCHACH *M:SUM C* ADJUSTMENT SCALE FOR HIGH AVERAGE INTELLIGENCE ADULTS

Rank[a]	Ratios										
7				3:2 3:4	4:3 4:5	5:4 5:6	6:5 6:7	7:6 7:8 7:5	8:7 8:9 8:6	9:8 9:10 9:7	10:9 10:8
6				3:3 3:4	4:4 4:2	5:5 5:3	6:6 6:4	7:7 7:9 7:4	8:8 8:10	9:9 9:6	10:10 10:7
5			2:1 2:2 2:3	3:5	4:6	5:7	6:8		8:5		10:6
4		1:1 1:3 1:2	2:4 2:0	3:1 3:6	4:7 4:8 4:1	5:8 5:2	6:3 6:2 6:10 6:9	7:3 7:10	8:4	9:4 9:5	10:5
3	0:1 0:2	1:4 1:0 1:5	2:5 2:6	3:0 3:7 3:8 3:9	4:0 4:9 4:10	5:10 5:1 5:0 5:9	6:1 6:0	7:2 7:1 7:0	8:2 8:3	9:2	10:3 10:4
2	0:3 0:4 0:5 0:6	1:6 1:7 1:8	2:7 2:8 2:9	3:10					8:0 8:1	9:1 9:0	10:2 10:1 10:0
1	0:7 0:8 0:9 0:10+	1:9 1:10+	2:10+								11:0

[a] High rank signifies high adjustment. See accompanying text for principles of construction.

quired for significance as the number of comparisons is increased. This suggestion was not adopted directly in the present research, but was kept, as it were, in the background. If *P* is raised to .001 rather than the usual .01 or less exacting standard, only two of the significance tests on personality measures in this study reach significance, *S* proportion to *R* on the Rorschach, and TAT Length of Stories. (Cf. Table 7.)

Before investigating thoroughly what others had done in making comparisons on multiscore tests, the writer had planned his statistics so that raw scores could be transformed into percentage scores or proportions for comparisons. He reasoned that if Rorschach *R*'s differed significantly (as they did differ in his data, it turned out), any further comparison of raw scores would merely be a function of the total responsiveness of the two groups.

On the Rosenzweig, raw scores are transformed immediately into percentages; but these both mean the same anyway, for there can be just twenty-four total scores, except in the rare case of unscorable responses.

It was felt that the same argument would hold for the TAT as for Rorschach; therefore, percentages or proportions were compared.

Small denominators in ratios may be unreliable, so Adjustment Scales were constructed (see Tables 4 and 5, e.g.) which did not depend on working out fractions. Ratios were transformed into

TABLE 5

RORSCHACH $FC:CF+C$ ADJUSTMENT SCALE FOR HIGH AVERAGE INTELLIGENCE ADULTS

Rank[a]	Ratios										
7								7:0	8:0	9:0	10:0
6					4:0	5:0	6:0 6:1	7:1	8:1	9:1 9:2	10:1 10:2
5		1:0	2:0	3:0	4:1	5:1	6:2	7:2	8:2	9:3	10:3 10:4
4		1:1	2:1 2:2 2:3	3:1 3:2 3:3 3:4	4:2 4:3 4:4 4:5	5:2 5:3 5:4 5:5 5:6	6:3 6:4 6:5 6:6 6:7	7:3 7:4 7:5 7:6 7:7 7:8	8:3 8:4 8:5 8:6 8:7 8:8 8:9	9:4 9:5 9:6 9:7 9:8 9:9 9:10	10:5 10:6 10:7 10:8 10:9 10:10
3		1:2 1:3	2:4 2:5 2:6	3:5 3:6 3:7	4:6 4:7	5:7 5:8 5:9	6:8 6:8	7:9 7:10	8:10		
2	0:1 0:2 0:3	1:4 1:5 1:6	2:7 2:8	3:8	4:8 4:9	5:10	6:10				
1	0:4+ 0:0	1:7+	2:9+	3:9+	4:10+						

[a] High rank signifies high adjustment. See accompanying text for principles of construction.

fractions only for making intragroup comparisons of rank orders on the two introversive-extratensive ratios yielded by the Rorschach, $M : Sum\ C$ and $(FM + m) : (Fc + c + C')$.

In working out the $M : Sum\ C$ and $FC : CF + C$ Adjustment Scales (Tables 4 and 5), ratios which were equal numerically were considered in part as equal psychologically, contrary to a warning by Cronbach. An examination of the $M : Sum\ C$ Scale (Table 4) reveals this fact patently in Rank 6 where ratios from 3:3 to 10:10 are grouped together. Whatever the meaning psychologically, however, the distributions on these two scales in both subject groups turned out to be fairly normal, though platykurtic.

Analysis of variance was discarded for comparisons because, in part, the experimental design was not planned as an analysis-of-variance experiment, but also because the data cannot be handled by analysis of variance. The many scores on each test cannot be considered as different scores from different individuals, or even as different scores from the same individuals on the *same* test, but as different (though related) scores on *different* (though related) tests.

After consideration of other tests of significance and consultation with statistical authorities, simple t tests were chosen to make comparisons between the two groups of persons.

With the Rosenzweig, since the same number of total responses is found in all cases, there is probability of correlations among percentage columns; hence, correlation coefficients were obtained between the Obstacle-Dominance, Ego-Defensive, and Need-Persistence columns for Conservatives. The *O-D* column is not significantly correlated with the others. But the *E-D* and *N-P* columns correlated with an r of $-.65$, which is significant at the 1 per cent level.

The comparison of *E-D* and $\underline{E + I}$ means of the two groups showed a significant t ratio (2 per cent level). Since $\underline{E + I}$ is a combination of scores taken out of the *E* and *I* scores, and specifically from the *E-D* column, the writer was not sure how to handle this combination set of scores. He decided that if there was not a significant correlation between *E-D* and $\underline{E + I}$, no dependent relation would be shown, and a t test could be safely interpreted for $\underline{E + I}$. As it is, however, the coefficient of correlation, .40, is significant at the 5 per cent level. So in the final analysis the $\underline{E + I}$ comparison was discarded.

A question may be asked here (or anywhere), "Are the distributions of percentage scores on the various tests relatively normal?" The answer is, "Yes, for the most part." In only a few cases the distributions were skewed considerably. But both Conservatives and Liberals had skewed columns, and in the same direction. On the basis of practically normal distribution of means of skewed distributions like these, it was decided that t tests could be utilized when skewness was practically the same in both curves to be compared.

A number of columns of both TAT and Rorschach scores contained too few scores to constitute a distribution in which t could be used on percentages of R. Therefore, when the scores that were present were scattered among a number of individuals and no one individual contributed an undue weight to the proportion, proportions of group scores to total responses were compared. The hypothesis to be tested was: The sum of Conservative or Liberal scores on each variable is in the same proportion to the combined Conservative and Liberal scores on that variable as the total number of responses of Conservatives or Liberals on the entire test is to the combined responses of both groups on the test. If the proportions differed significantly, the hypothesis was rejected.

Besides the foregoing, one minor comparison was made, that between rank-order correlation coefficients on the two

TABLE 6

SIGNIFICANT CORRELATIONS SHOWING RELATIONS BETWEEN TESTS AND WITHIN TESTS

Test or Factor	*df*	r or *rho*	Significance Level[a]
Salvation Opinionnaire and Ferguson (Subjects and eliminees)[b]	38	.73	1%
Salvation Opinionnaire and Ferguson (Subjects only)[b]	28	.77	1%
Rosenzweig *E-D* and *N-P* (Conservatives only)[b]	28	−.65	1%
Rosenzweig *E-D* and $\underline{E+I}$ (Conservatives only)	28	.40	5%
Rorschach (M:*Sum* C) and ($FM+m$:$Fc+c+C'$) (Conservatives only)[c]	28	.39	5%

[a] Refers, of course, only to probability level, not degree of association.

[b] Pearson product-moment r.

[c] Rank-order correlation between two Introversive-Extratensive ratios.

introversive-extratensive ratios mentioned previously. The question to be answered was whether or not there was any higher correlation between these two supposed measures of introversion-extratensiveness in one group than in the other. The coefficients were transformed into Fisher's z function in order to utilize a t test, which depends on the assumption of a normal distribution of sampling means or other statistics to be compared. It can be seen from Table 6 that the true correlation for Conservatives may be significant. That for Liberals was not significant; but a t test revealed no significant difference between the Conservative and Liberal coefficients.

Several original adjustment and balance scales were constructed in order to make further comparisons which cannot be made in simple percentages and proportions. These made use of Rorschach raw scores in their respective ratios. The first of these scales simply listed $M : \mathit{Sum}\ C$ ratios as introversive, extratensive, or "no trend" on the basis of Rorschach theory. Even though additional M, FM, and m were not included, as Klopfer (11, p. 253) suggests they might be, both groups were very much overweighted in the introversive direction, Conservatives' frequencies being 22 introversive, 1 "no trend," and 7 extratensive, and Liberals' frequencies 24, 5, and 1 respectively.

The other ratio indicating supposedly a person's *Erlebnistyp*, $(FM + m) : (Fc + c + C')$, yielded the same results as $M : \mathit{Sum}\ C$. Conservatives' ratios scored 24 introversive, 1 "no trend," and 5 extratensive, and Liberals' 23, 2, and 5 respectively. If the two groups of this study manifesting such strong introversial trends on the I-E ratios are at all representative of the population at large, questions may be asked either about the adequacy of the ratios or the usual assumptions about the distribution of introversion and extratensiveness.

In conjunction with the above indicators of *Erlebnistyp*, the third major indicator suggested by Klopfer, the percentage of responses to the last three cards came the closest to theoretical expectations. Conservatives averaged 41.32 per cent with a standard deviation of 11.55; Liberals averaged 39.28 per cent, standard deviation 11.05.

$M : \mathit{Sum}\ C$ ratios were compared for the two groups by means of frequency distributions on the Adjustment Scale represented in Table 4. Prior to examination of the data from subjects studied, ratios within reasonable limits of $M : \mathit{Sum}\ C$ were grouped into seven classes on the basis of the following principles.

1. A slight unbalance either way represents a more favorable adjustment than an exact balance.
2. A greater unbalance of $\mathit{Sum}\ C$, in the smaller ratios especially, means a poorer adjustment than the same for M.
3. The quality of M and the presence of C are disregarded.
4. For above-average intelligence adults, at least three M's should be produced for best adjustment.
5. Where a ratio has a fraction, .5, the figure is raised to the next higher whole number for matching to the Scale.

Whether the Scale has any psychological meaning is one question; whether on two distributions of ratios it appears to hypothetically correct, the question is answered affirmatively. A chi-square test of the hypothesis of normal distribution (Guilford, 10, p. 284f.) produced a chi square for Conservatives of .98. The ends of the theoretical and observed distributions were combined, making only three sets of discrepancies, but the number of degrees of freedom was judged to be four (seven class intervals). Even if there were only one degree of freedom, however, .98 would not be significant. Other chi squares were not worked out for Liberals or for either group on the other Adjustment Scale below, since the forms of their distributions approximated the one already tested.

Table 5 presents the ratios of *FC* : *CF* + *C* divided into seven classes, a set of rankings set up before the data from the research groups was examined. This Scale is not quite so successful from the standpoint of approximating normality as the *M* : *Sum C* Adjustment Scale. It may, however, represent more psychological meaning than the other, since the upper levels of optimal adjustment include fewer persons than on the *M* : *Sum C Scale*. The principles of construction of the scale are the following.

1. A balance toward *FC* is considered more favorable to adjustment.
2. The presence of *C per se* is disregarded.

Besides the above Scales, several Adjustment Dichotomies were constructed for comparison purposes. An $(FM + m)$: $(FC + c + C')$ Balance Adjustment Dichotomy was worked out on the following simple principles.

1. The more evenly balanced ratios are regarded as more favorable to adjustment (3:2, 2:3, etc.).
2. The absence of one or the other member of the ratio is not considered unbalanced if the ratio is small.

Apparently, this balance scale is satisfactory theoretically, for both groups produced a large majority of "balanced" individuals, 23 to 7 in each group, meeting expectations in a relatively normal population.

The two other dichotomies worked out to make comparisons between the groups were not so satisfactory from a statistical standpoint. A *W* : *M* Balance was constructed on the following principles.

1. At least a 2:1, and better, a 3:1 ratio is held necessary for balance.
2. At least three *M's* are considered necessary for balance.

The ratios divided the Conservatives into sixteen "balanced" and fourteen "unbalanced" persons and the Liberals into nineteen "balanced" and eleven "unbalanced." Additional factors such as the quality of *W* and *M* may be needed to make this Balance scale more meaningful.

A similar criticism of the statistical value of the $M : FM + m$ Balance can be made. This is a dichotomy in which ratios are divided into balanced and unbalanced according to the principles below.

1. At least two *M's* and a ratio of 1:1 or at the most 2:3 are held necessary for good balance.
2. A preponderance of *M* is more indicative of adjustment than one of $FM+m$.
3. When larger numbers of *M* and $FM+m$ are produced, an unbalance in favor of $FM+m$ is not destructive of balance.
4. The presence of larger amounts of *m* compared to *FM* has not been regarded in making divisions.
5. The quality of the responses is disregarded.

The actual divisions into "balanced" and "unbalanced" ratios happened to be exactly the same as with the *W* : *M* scale, sixteen and fourteen for Conservatives, and nineteen and eleven for Liberals.

"Gestalt limits" presented by both Klopfer and Mons for $W : D : d : Dd + S$ were supposed to be indicated by ranks ranging from 1 to 4, the latter representing the best adjustment, with an individual assigned a rank depending on how close he fell within the limits set forth in Klopfer's "Manner of Approach." It was found that *no one in either group* fell exactly inside the limits, and most persons fell far outside. Again, either the two groups are entirely apart from the average or normal, or new norms are needed. The writer believes from other experience as well as this that serious questions can be raised about these so-called "gestalt limits."

The various scales which have been described are offered not only as the means by which the data of this research were ordered in part, but also as simpler models of what seems possible on a more complex plane. Rorschach experts have much to say about configurations and interaction of one score with others. But it seems that by grouping scores together that appear to go together and examining large numbers of normal records an empirical set of scales could be derived that would not depend only on how many different factors an intelligent interpreter can keep in mind.

TABLE 7

SUMMARY OF SIGNIFICANT DIFFERENCES BETWEEN RELIGIOUS CONSERVATIVES AND LIBERALS ON PERSONALITY TESTS

Test Factor	Group having higher Mean Score, Mean Percentage, or Proportion	t	Significance Level
Rosenzweig *E-D*%	Conservatives	2.45	2%
Rosenzweig *E+I*%	Conservatives	2.64	2%
Rorschach *R*	Liberals	2.92	1%[a]
Rorschach *F+*:*R*	Liberals	2.58	2%
Rorschach *FK*:*R*	Liberals	2.14	5%
Rorschach *S*:*R*	Liberals	4.13	1%
Rorschach *P*%	Conservatives	2.77	1%
TAT Length of Stories	Liberals	5.28	1%
TAT *n Dep*, *-n Dep* (*I*)[b]%	Conservatives	2.64	2%
TAT *n Dep*, *-n Dep* (*I*):*R*	Conservatives	2.01	5%

[a] 2% with Adjusted Liberal mean.

[b] This refers to Column 9 defined in text.

VI. CONCLUSIONS

A. QUANTITATIVE COMPARISONS

Out of 52 t ratios computed, nine significant differences were found between percentage means, proportions, and two of the raw scores which were compared (Rorschach *R* and TAT Length of Stories). These are summarized in Table 7. (The last line of the table is a variant of the ninth.) In order to evaluate the summary correctly two facts must be kept in view. First, at the 5 per cent level of significance one would expect two or three out of 52 t tests to be significant just by chance. And secondly, 43 out of 52 tests were *not* significant.

Relating the quantitative results to the hypotheses set up to guide research, one finds that the majority of the hypotheses must be rejected.

Hypothesis 1. *Religious liberals are more "emotionally mature" than religious conservatives.* Considering "emotional maturity" as a summation of individual aspects of "maturity" represented by scores on the personality tests, it was concluded that this hypothesis had to be rejected, or accepted only as applicable to limited areas. Four of the nine scores (Table 7), *E-D*, *F+*, *FK*, and *S*, were specifically hypothesized as testing maturity. The first three are in the direction of the hypothesis. Therefore, it might be said that in the areas of perceptual keenness and insight Liberals appear to be more fully developed (mature) than Conservatives. If the *E-D* t ratio can be trusted, Liberals also tend to be less ego-defensive.

The other scores in Table 5 were not specifically set up for testing Hypothesis 1. It is not accepted scientific practice to base conclusions *ad hoc* on comparisons made after the data are inspected (though the principle is not observed in many instances). Therefore, these differences can be used as suggestive hypotheses for future research, but not directly as means of proof or disproof of the present proposition.

After all the significant differences are reviewed, however, and even if all of them should be admitted (if p should be set at .001 for significance, only two could be admitted), the fact stated above must be remembered: that on 42 other scores, most of which, if not all of which, are presumably independent, there are no significant differences. The majority of these 42 were specifically hypothesized in connection with the "emotional maturity" hypothesis, so must be held to be germane to its acceptance or rejection.

Hypothesis 2. *Rigidity of personality structure and conservative religious attitudes are positively correlated.* If Rosenzweig *E-D* and $\overline{E + I}$ were in columns uncorrelated with each other or with other columns, this second proposition might be accepted in part, since Conservatives could be shown to possess comparatively greater defensiveness or rigidity as measured on the Rosenzweig. But because correlations were found between *E-D* and other columns, linear relations (on which significance tests depend) most likely do not hold; therefore, the t ratios are suspect. The one major score on which a difference was expected to occur, *F*% on the Rorschach, did not show significant mean differences.

Hypothesis 3. *Different types of emotional control are found in both groups.* It was anticipated that the Rorschach in particular would show that the Conservatives employ more intellectualized, formal types of control, while Liberals employ more emotional, inner types of control. The fact is, as tested by this study, no significant differences appear between the groups. Therefore, this hypothesis is rejected.

Hypothesis 4. *Conservative individuals are more guilt-ridden than liberals.* It was specifically stated prior to making comparisons that Conservatives would show higher indices in reference to guilt than Liberals on Rosenzweig *E-D,* TAT *n Agg (I)* and *Guilt,* and in corollary fashion would show more inadequacy and insecurity feelings as measured by TAT *Adeq* and *Anx* (and their negatives). Since the t ratio on the first score has possibly an ambiguous meaning, and none of the other t ratios approaches significance, Hypothesis 4 is not accepted.

Hypothesis 5. *Liberals tend to gloss over aggression-provoking situations more than conservatives. M* on the Rosenzweig, implying impunitive reactions to aggression, was thought above other scores to tend to measure a glossing over of aggression-provoking situations in connection with this hypothesis. No real difference obtains between the two groups according to this test, so the hypothesis is rejected.

Hypothesis 6. *A greater need for dependence is found in the conservative.* This is the one proposition among six which is accepted, though with reservations. Column 9 of the TAT, including scores *n Dep, −n Dep (I),* has been shown to be somewhat suspect, in that on twenty cases apiece (the group in which the effects of possible scoring errors are removed), the significant t ratio disappears. It is not certain, however, that scoring errors are responsible.

In no other comparison of twenty cases apiece were the *t*-test results obtained on thirty cases changed. There is in Column 9 an overweighting of scores in the first ten cases among Conservatives. This condition could arise from scoring errors; but since in no other case did such overweighting occur from scoring errors, it seems reasonable to believe that the higher TAT scores (Column 9) in the first ten cases reflect the actual condition. So when they were eliminated, the other twenty scores did not truly represent the total distribution of Conservatives.

Also hypothesized, as a negative instance, was the Rorschach score *S*. Although it has other meanings, it is a sign of independence. The 1 per cent level of significance (in fact, the .001 level) favors Liberals in independence.

A further consideration that cannot be disregarded is the difference between *P%* means (1 per cent level). Although this score was not specifically hypothesized for the dependence-independence continuum, the fact that a real difference of means exists between the groups on *P%* adds no reason to reject the hypothesis.

On the opposite side of the ledger is Rorschach *A%*. It was strongly suspected that Conservatives would be much higher in this category. The fact that the obtained *t* ratio could easily occur by chance tends to militate against the acceptance of Hypothesis 6.

The writer must admit that the hypotheses he set up originally are not all clear-cut and admissible of unequivocal answers, as good hypotheses should be. They seemed to be satisfactory when first proposed, but the research showed how "fuzzy" several of them are. Future research along this line will have to propound more single-minded hypotheses.

B. Qualitative Comparisons

When the two groups under study are compared qualitatively, the same results approximately are yielded as in the quantitative comparisons. At the conclusion of testing all subjects, the writer had only one clearly defined opinion, that wide differences existed among individuals in the groups; but he had no idea whether these differences represented groups as such. In processing results the scores were so many and the scoring so complicated that it was impossible to tell at any time which group had more "balanced" scores. This inability to tell which way balance lay continued up to the very last comparison of all significance tests.

If further treatment of the writer's data or other future research with more refined methods should yield different conclusions, the writer would not deny that more adequate methods can produce different results. He feels that though his study has its values, it also has limitations.

The major qualification to be kept in mind in interpreting results is that only 490 out of 1035 persons returned the Salvation Opinionnaire. Such a situation introduces a selective factor which may make for nontypical samples. This response factor operated to a greater extent among Liberals than among Conservatives, for twenty-five out of thirty Liberals were selected from letter-distributed returns and only five out of thirty Conservatives were so selected. Group distribution yielded virtually 100 per cent returns, so the selective factor under consideration did not affect Conservatives to an appreciable degree.

One mitigating circumstance may indicate that the Liberal sample is representative of religious Liberals as a whole. Wonderlic (intelligence) scores revealed that Liberals were more intelligent than Conservatives (1 per cent level), when comparisons were made involving both subjects *and* eliminees (those with scores below 20 on the Wonderlic or inconsistent scores on the two religious tests). This finding is in keeping with previous research.

In the meantime it may be stated that if these samples are representative of their respective religious populations, and if the tests actually measure what they are supposed to measure, and if the writer's reliability as examiner and scorer is sufficient, the religious populations designated as "Liberal" and "Conservative" appear to be one in matters of personality and of maturity, with only one admissible exception found in this study, that of greater dependency among conservatives.

VII. SUMMARY

1. The Salvation Opinionnaire (19) was sent to or administered directly to 1035 persons, of whom 490 returned the Opinionnaire and information sheet.

2. Sixty persons (30 persons in each group) were chosen whose scores on the Opinionnaire and Ferguson's Primary Social Attitude No. 1 (Religionism) indicated representativeness of religious liberalism or conservatism, and whose Wonderlic Personnel Test (intelligence) scores were 20 or above. Other means of selection were employed to assure comparability on nonexperimental factors.

3. Rosenzweig's Picture-Frustration Study, the Rorschach, and the Thematic Apperception Test were administered to the sixty subjects.

4. Selected variables relating to "emotional maturity" were compared by means of t tests of percentage, proportion, or ratio-frequency scores.

5. It was found that nine out of fifty-two significance tests showed statistically significant or very significant differences between means or proportions.

6. On the whole it was found that the two religious-attitude groups could be considered samples from the same population as regards personality factors, with the likely exception of dependency feelings in which a greater need is shown by the religious conservatives.

No specific meaning has been attached to the greater Liberal responsiveness on the Rorschach and TAT, since there seems to be no agreement on specific significance of the amount of responsiveness, except for pathological rejections and consequent constriction of records.

What meaning, in terms of explanation, greater dependency on the part of Conservatives has can only be conjectured at this time. It may mean, on the analogy of the experience of others who have been investigated in this area of dependence, that such dependence has been fostered from early childhood. Or it may mean that the type of religious training which Conservatives have received tends to promote feelings of dependence. Of course, some combination of these two aspects may be the actual situation. The writer is inclined to believe it is more probable that those with similar types of dependency needs, arising out of infancy and early childhood training, would band together for mutual support.

BIBLIOGRAPHY

1. Beck, S. J. *Rorschach's test.* Vols. I & II. New York: Grune & Stratton, 1944 & 1946.
2. Bell, John Elderkin. *Projective techniques: a dynamic approach to the study of personality.* New York: Longmans, Green & Co., 1948.
3. Bernard, Jack. The Rosenzweig Picture-Frustration Study: I. Norms, reliability, and statistical evaluation. II. Interpretation. *J. Psychol.,* 1949, **28**, 325-332, 333-343.
4. Cronbach, Lee J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.,* 1949, **46**, 393-429.
5. Cronbach, Lee J. Statistical methods for multi-score tests. *J. clin. Psychol.,* 1950, **6**, 21-26.
6. Dreger, Ralph Mason. Some personality correlates of religious attitudes as determined by projective techniques. Unpublished Ph.D. dissertation, Univer. Southern California, 1950.
7. Ferguson, L. W. Primary social attitudes. *J. Psychol.,* 1939, **8**, 217-223.
8. Ferguson, L. W. The stability of the primary social attitudes: I. Religionism and II. Humanitarianism. *J. Psychol.,* 1941, **12**, 283-288.
9. Guetzkow, Harold. Unitizing and categorizing problems in coding qualitative data. *J. clin. Psychol.,* 1950, **6**, 47-57.
10. Guilford, J. P. *Fundamental statistics in psychology and education.* Second edition. New York: McGraw-Hill, 1950.
11. Klopfer, Bruno, & Kelley, D. G. *The Rorschach technique.* New York: World Book Co., 1946.
12. Mons, W. *Principles and practice of the Rorschach Personality Test.* Philadelphia: J. B. Lippincott, c. 1949.
13. Murray, H. A. (Ed.). *Explorations in personality.* New York: Oxford Univer. Press, 1938.
14. Shrout, Irwin E. A study of the relationship between emotional maturity and religious convictions and attitudes toward God of 300 college students. Unpublished master's thesis, Univer. Southern California, 1946.
15. Stein, M. I. *The Thematic Apperception Test: an introductory manual for its clinical use with adult males.* Cambridge, Mass.: Addison-Wesley Press, 1948.
16. Symington, Thomas A. *Religious liberals and conservatives: a comparison of those who are liberal in their religious thinking and those who are conservative.* New York: Teachers College, Columbia University, Contributions to Education, No. 640, 1935.
17. Tomkins, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.
18. Webb, W. B. & Lemmon, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.,* 1950, **47**, 130-136.
19. Wendland, Leonard V. The Salvation Opinionnaire. Unpublished paper, Graduate School of Religion, Univer. Southern California, 1949.
20. Wesley, S. Medford. Thematic Apperception Test analysis guide. Tentative draft. Univer. Southern California, 1948.
21. Willoughby, R. R. Emotional maturity of some religious attitudes. *J. soc. Psychol.,* 1930, **1**, 532-536.
22. Wonderlic, E. F. *Wonderlic Personnel Test manual.* Glencoe, Ill.: Privately published, 1945.
23. Woodward, Luther Ellis. *Relations of religious training and life patterns to the adult religious life: a study of the relative significance of religious training and influence and of certain emotional and behavior patterns for the adult religious life.* New York: Teachers College, Columbia University, Contributions to Education, No. 527, 1932.

VOLUME 66
NUMBER 4

WHOLE NO. 336
1952

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Personal Security as Related to Station in Life

By

ANDIE L. KNUTSON
Federal Security Agency
Public Health Service
Washington, D.C.

Accepted for publication July 2, 1951

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

ACKNOWLEDGEMENTS

THE WRITER desires to express his deep appreciation to all those who have helped in the completion of this study.

A fellowship grant from the Social Science Research Council made my graduate work possible. The Office of Public Opinion Research paid the expenses incurred in printing and administering the questionnaires used. I acknowledge this assistance most gratefully.

I wish to thank several hundred employees of a veterans' mental hospital, many patients of this hospital, employees of a welfare agency, and students of Princeton University for contributing by patiently answering questions about their personal feelings.

My thanks go also to the members of my advisory committee, Professors Hadley Cantril, Carroll C. Pratt, and Gustave M. Gilbert for reading the manuscript and offering helpful criticisms; and to Professors Herbert S. Langfeld, Leo P. Crespi, and George Horsley Smith of Princeton University, and Dr. Bert D. Schwartz, Clinical Psychologist, Newark Clinical Group, for suggestions and other assistance. Elizabeth V. Deyo has helped in a number of ways.

The opportunity of carrying out this study under Professor Hadley Cantril has been a privilege for which I shall always be grateful.

This work could neither have been undertaken nor completed without the fullest cooperation, encouragement, and faith of my wife, Ruth Knutson.

ANDIE L. KNUTSON

PERSONAL SECURITY AS RELATED TO STATION IN LIFE[1]

I. INTRODUCTION 1
II. METHOD 3
III. VALIDATION 7
IV. THE SAMPLE AND DISTRIBUTION OF QUESTIONNAIRES 9
V. THE AREAS OF SECURITY AND INSECURITY 11
VI. THE GENERAL LEVEL OF SECURITY 17
VII. SUMMARY AND CONCLUSIONS 24
REFERENCES 25
APPENDIX A 26
APPENDIX B 28

I. INTRODUCTION

THE concept of personal security is frequently employed by social scientists, but many of those who find the concept useful are not clear as to what they mean by the term. What some writers refer to as security or insecurity is defined by others as adjustment or maladjustment, high or low morale, and the like. The general purpose of this study was to investigate the nature of personal security to determine what it means psychologically, and to gain some insights about correlates of personal security. This involved several steps:

1. An investigation was made of the definitions and usages given to the concept of personal security by many psychologists and other social scientists who have invoked this concept in recent years. In spite of wide variation in systematic treatment and terminology, the following nucleus of agreement was found:

> Personal security is a "dynamic" concept. People seek security within their various areas of striving: by achieving it, they become secure in some areas; by failing to achieve it, they become insecure in other areas. The areas of personal striving are interrelated and overlapping, and some may be of greater importance than others to the security of the individual. Feelings of security or insecurity may tend to become generalized throughout the personality structure, or they may be projected from one area of striving into others. Furthermore, the biological structure or physical or social environment of an individual, by limiting or influencing the character and direction of his motivations, may have a direct bearing upon the nature of his securities. In view of these conditions or possibilities, there appears to be a general level or status of security, as well as levels of security within the individual areas of striving.
>
> A person's feelings of security or insecurity within any area of striving involve his own subjective evaluation of his success, satisfaction, and surety or confidence with respect to the carrying out of his purposes in past and present situations and group relationships, also, his expectations, hopes, fears, or uncertainties with respect to the carrying out of his purposes and aspirations in future situations and group relationships.
>
> Feelings of security within any functional area vary on a continuum, so that security status within any area of striving or within any situation may be considered a matter of degree rather than a matter of absolute.

2. The next objective of the present investigation was to develop a questionnaire based upon this concept of personal security, and to administer this

---

[1] The opinions or assertions contained in this paper are those of the author and are not to be construed as official or as reflecting the views of the U. S. Public Health Service.

questionnaire to groups that are ostensibly secure and to those who have been classified by psychiatrists as "insecure." This was intended to provide a test of the validity of the concept.

3. It was then planned to administer the validated personal-security questionnaire to a sample of adults varying in income and occupation for the purpose of comparing the security of different income and occupational groups. This was designed to test the thesis that otherwise normal persons who lack economic security tend to feel insecure in the more developmental areas of striving, such as belongingness, sex, personal appearance, cultural activities, etc. It would also test the hypothesis that patterns of security vary within a population according to income, occupation, education, etc.—what might be termed their station in life.

II. METHOD

It was decided to investigate security feelings by means of a questionnaire that could be used in a group situation as a pencil and paper test, or, with some variation, in a direct-interview situation as a public opinion schedule. The questions were selected with an eye to satisfying the criteria of opinion research.

Many of the limitations of paper and pencil tests of personality have been pointed out in the literature, and these draw attention to the need for interpreting with caution and reservation findings based on such methods. It has been emphasized that individuals may not know or understand what motivates them, and that for this reason we may never be able to get at basic motives by asking questions. Adequate alternatives to fixed questions may not be offered because they may not even be known. When intimate or emotional matters are touched upon, blocking may occur. Conscious and subconscious cheating is likely to occur when the ego is involved, etc.

These limitations, common to all personality questionnaires, have been recognized, but in view of the subjective nature of security feelings, no alternative to questioning as a means of studying security seems possible. Truman Kelley has pointed out, "If facts of a unique nature, known only to specific individuals, are desired, then an oral or written questionnaire addressed to the persons knowing these facts is the only possible method of approach" (8, p. 35).

In developing our questionnaire, a selection was first made of certain major areas of striving within which people may be secure or insecure. Then a ranking method was developed for obtaining a rough grouping of these areas in terms of their importance to individuals. Finally, a questionnaire was designed to get at personal feelings of security within each of these major areas.

SELECTION OF THE AREAS OF STRIVING

In choosing the areas of striving within which to measure security feelings, an attempt was made to select broad areas which would apply generally to the majority of the adult population. It was apparent that the areas could not be specific enough to include the multitude of unique motives we each have and still be useful from the practical point of view of measuring large groups of people. The areas used in this study were chosen because they appeared on the basis of former reports and the preliminary investigation to be most pertinent.

As the first step toward determining the areas of striving in this study, fifteen mature individuals were interviewed using the open-interview method. The prime purpose of these interviews was to obtain some insight into what frames of reference people use in their discussion of security insofar as it applies to them personally. The key questions used in this preliminary investigation were: "What do people need to feel secure?" "What would you need to feel secure?" and "What would make you feel most insecure?" Numerous nondirective probes were used, and care was taken to avoid suggestion.

A group of twenty undergraduate students in psychology was then asked to write out an answer to the question, "What would you need to feel secure?" They were asked not to identify their papers in any way, and to feel free to write frankly and honestly.

The responses obtained in these preliminary investigations varied considerably, giving evidence of the unique nature of individual feelings of security. In spite of this variation, however, three general areas of striving were mentioned as the frame of reference very frequently. These were the various aspects of economic security, belongingness, and mental or intellectual capacities. Often these seemed to overlap; at times one appeared to be regarded as a stepping stone to another.

A group questionnaire was then constructed and given to a group of thirty undergraduates in psychology who were likewise instructed not to sign their names and to feel free to answer the questions honestly and frankly. In this questionnaire the subjects were asked to rank a number of areas of striving in terms of importance. The method was similar to that previously used by Hyman in his study of social status (7). This method was not completely successful because many of the subjects appeared to have difficulty in selecting between some of the alternatives. It was apparent that if these finer discriminations were difficult for college students in a classroom, they would be far too difficult for individuals with less education and test experience. A coarser grouping seemed desirable.

In the second part of this questionnaire, the subjects were asked to rate each of the areas of striving on a five-point scale in terms of their present feeling of satisfaction; then they were asked to rate them again on a similar scale in terms of the level of satisfaction they hoped to achieve. This attempt to obtain a measure of the level of security within each of the areas showed some promise, but it was abandoned as too complex for people with limited test experience. Furthermore, individual frames of reference with respect to the use of rating scales seemed to influence the nature of the final scores which involved a comparison of the two ratings.

In addition to these ranking and rating questions, this questionnaire again included open questions designed to draw out further information about the areas of security and insecurity.

On the basis of the review of previous studies of value and ego-involvements and the preliminary investigations, it was decided to study personal security relative to ten areas of striving: *cultural, social standing, intellectual, economic, sexual, appearances, belongingness, religious, health,* and *work satisfaction.* Concrete formulations of the areas were developed through pretesting, as follows:

Having a chance to develop some special talent, interest, or hobby.

Having a good social standing and the right kind of people as friends.

Having a good ability to understand things, good education, and general information.

Being sure of a job or steady income with good living conditions.

Being satisfied in your relations with members of the opposite sex.

Having good looks and appearances, good build, height, etc.

Being liked and respected, and feeling you "belong."

Having a good religious understanding and faith.

Being in good physical and mental health.

Being satisfied with your work or job.

These areas of striving are not independent and distinct. A person very insecure in the subjective economic area, for example, would almost certainly feel insecure insofar as his social status and his possibilities for developing his mental and intellectual capacities are concerned. Then, too, as has been suggested, insecurity feelings may be generalized beyond the limits of an area, and insecurity in one area may be projected into other areas, etc.

The advantages of using such an area approach offset these difficulties. The areas served as a framework in developing the personal-security questionnaire in such a way that a weighting of the scores in any single direction could be avoided. Not enough is yet known about the relative importance of different values in terms of motivation to justify an overemphasis in any particular direction. In addition to providing a framework for the questionnaire, the area approach could be used to determine what people feel to be most important—what they say they are striving for. Thus, the questionnaire could be designed to obtain a measure of the general level of security and at the same time a description of the most important areas of striving.

THE RANKING METHOD

Several methods were considered as possibilities for obtaining a ranking of the areas of striving in terms of their importance to the individual. In the preliminary investigations, subjects were asked to rank all the values in terms of importance. This proved too difficult for many subjects, owing to the equal importance given to some areas and the interrelationships existing between areas. It was, therefore, decided to obtain a rougher three-way grouping.

Three ways of obtaining this ranking were tried out in the first complete trial of the Personal Security Inventory. At the beginning of the questionnaire, the concrete formulations of the areas were presented in a list and the subjects were asked:

> 1. a. "Which *three* of these do you think are most important?
> That is, which three do you care most about?"
> (Make three checks in the first column.)
> b. "Which *three* do you care least about?"
> (Make three checks in the second column.)

In the second presentation, the same concrete formulations of the areas were presented in a different order, and the subjects were asked:

> 20. a. "Which *three* of these worry you the least?"
> (Make three checks in the first column.)
> b. "Which *three* of these worry you the most?"
> (Make three checks in the second column.)

Then, at the end of the Inventory, the subjects were asked:

> 42. a. "What do you think people need to feel most secure?"
> b. "What else?"
> c. "What do you think makes people feel most insecure?"
> d. "What else?"

In the validation test of the Inventory, using 73 subjects, there was close agreement between the responses obtained by the different approaches. In nearly all cases, one of the three areas checked as important was also checked as an area worried about; and when responses to the open questions at the end of the Inventory were coded in terms of the ten formulated areas, one or more of the areas mentioned in the open questions was included among those mentioned in both of the ranking questions three-fourths of the time.

Since it was necessary to cut down the length of the Inventory so as to be able to obtain adequate descriptive data of the subjects at the same sitting, one of the presentations had to be dropped. There was no adequate means of judging the least valid of the approaches. The first question was retained as being the one least likely to elicit an emotional blocking. The final open questions were also retained as a check on the responses to the ranking question and in order to leave room for unique individual expression.

THE SECURITY QUESTIONNAIRE

The major areas of striving were used as a framework in developing the fixed alternative questions of the Personal Security Inventory, so that undue weighting of the questionnaire in any single direction was minimized. The approach was also considered useful for disclosing patterns of security and insecurity which might then be studied more fully.

For each of the areas, questions were designed to reveal the individual's own subjective evaluation of his success, satisfaction, and surety with respect to the carrying out of his purposes in past and present situations and group relationships, and, also, with respect to his expectations about carrying out his purposes and aspirations in future situations and group relationships. In accordance with the concept of security accepted for this investigation, an individual who felt unsuccessful, dissatisfied, and unsure of himself with respect to an area of striving, and who was doubtful or fearful of the future, was considered to be very insecure insofar as that area of striving was concerned.

With the exception of the subjective economic security and religious areas, four questions were asked relative to each major area. Eight items were included in the economic area to permit a better test of certain of the major hypotheses of the study. The ranking of the religious area provided a measure of belief but no attempt was made to measure religious security.

METHODS OF SCORING

The security questions in the Inventory were provided with three response possibilities (Satisfied, Dissatisfied, Don't know). Of these, the positive response (Satisfied) was scored *1* as "secure," and both the negative and the doubtful responses were scored *0* as "insecure." Since the directions were very emphatic about answering every question and requested a final review of the Inventory to be certain that every question was answered, failure to answer a specific question was also scored *0*.

The assumptions here are that if a person does not know whether he is satisfied or dissatisfied, he is not really very satisfied, and by and large the more insecure person is likely to fail to answer more questions than the secure person. An analysis of the doubtful and no answer responses for the questionnaires completed in the preliminary validation of the questionnaire bore out these assumptions.

Using this method of scoring, a total security score of 41 was possible; the highest obtained in the present investigation was one 40. It is possible that certain of the more discriminating items deserve a greater weighting than *1*, but there was little justification for assigning weights in advance.

FORMS A AND B OF THE INVENTORY

To minimize the influence of item order on responses, two forms of the Inventory were prepared. The only difference between these two forms is the order of the items on page one. Forms A and B were distributed in alternative order to respondents.

III. VALIDATION

The primary criticism leveled against paper and pencil personality questionnaires is that they have not been adequately validated against outside criteria. To meet this criticism, a thorough test of validity was made of the Personal Security Inventory using the preliminary form. This was followed by a second validation test with the final form of the Inventory.

PRELIMINARY VALIDATION STUDY

The preliminary questionnaire was administered in group situations to three groups of patients and one group of employees of a veterans' hospital. Members of all groups were white male war veterans; all could read and write. The groups used were as follows:

Neurotic group.—Fifteen male white war veterans who had been diagnosed by members of the staff of the hospital as "psychoneurotic." No extreme cases were included in the group; all were in an open ward and could come and go freely. They had been in the hospital less than six months. They were assumed to be very insecure.

Medical group.—Twenty white male war veterans who had been hospitalized and were receiving medical treatment for neurological injuries or difficulties. They were assumed to be fairly insecure.

Unclassified group.—Eighteen white male war veterans who had just been admitted to the hospital for guidance or treatment, and who had not yet been classified by members of the hospital staff.

Employee group.—Twenty white male war veterans who were employed in various positions in the same hospital. No member of this group had a history of psychiatric disorder, and the group was assumed to be the most secure of the four studied.

In addition to being white, male, and war veterans, the four groups were fairly similar in age and in weekly income, although there is some question as to the meaning of an income comparison insofar as hospitalized veterans are concerned.

The three groups of patients included a fairly equal proportion of married and single patients. A slightly larger proportion of the employees was married.

Considerably more of the patients in each of the three groups were Catholic than members of other faiths. This was not true of the employee group.

When a comparison of the mean security scores of the four groups was made, the neurotic group was found to differ significantly from all other groups at the 1.5 per cent level or lower. Differences between the means of the other groups were also in the expected direction[1] (see table below).

Group	*N*	*Mean*	*SE Mean*	*T Test for Neurotic vs. Other Groups*
Neurotic	15	13.73	2.33	—
Medical	20	21.00	1.75	2.55*
Unclassified	18	22.17	2.23	3.14
Employee	20	24.95	1.35	4.42

* Significant at the 1.5% level. The other values in this column indicate significance at the 1% level or lower.

Following the validation pretest, a detailed item analysis was made on the basis of which several questions were revised and a number of questions added in order to obtain sharper group discrimination. Tetrachoric correlations were made between the questions in each of the major areas of striving. The questions retained were those with the lowest intercorrelations with other questions relating to the same area of striving, but which were most effective in differentiating the neurotic group from the employee group.

---

[1] These findings are not directly comparable to data presented in Chapters V and VI, since the preliminary form of the Inventory was used.

FINAL VALIDATION STUDY

A final validation check was made of the Inventory using a larger number of psychoneurotic patients and several occupational groups at the same veterans' mental hospital. At this time the Inventory was found to discriminate significantly between the patient group and all other seven groups studied, and, also, between certain of the occupational groups themselves. These data are reported fully in Chapter VI.

It seems important to emphasize that these were group-validation studies. No claims are intended for the use of the Inventory in individual diagnosis. An argument might be made, however, for the face validity of the Inventory when individually administered, inasmuch as the responses involve a person's own subjective judgment of his feelings of satisfaction, status, surety, and expectation in the areas of striving covered by the Inventory. In other words, a person who ranks low on the Inventory when it is individually administered is classifying himself as insecure. The converse also holds, but the questions are all straightforward and frank, without concealed meanings or implications, so that a person may answer them falsely if he so desires.

IV. THE SAMPLE AND DISTRIBUTION OF QUESTIONNAIRES

Responses to the Personal Security Inventory were obtained from three major groups: 51 students in psychology at Princeton University; 34 psychoneurotic patients at a veterans' mental hospital; and 315 employees at the same veterans' hospital.[1]

Princeton Students

The 51 Princeton students were attending the summer session at the University. Many of these students were making up credits because of failures during the previous year. All were male, white, and under thirty years of age; nearly 9 out of 10 were single, and 8 out of 10 had served in the armed forces.

In family income the students as a group were considerably better off than any other sample studied. In relative income, however, the students did not differ much from the other groups. In fact, there was little difference in the distributions of the four groups in relative income, in spite of wide differences in actual family income and occupational status.

Another striking difference between the student and other groups was the higher educational and occupational status of the fathers of the students. The students also differed from other groups in that a much higher proportion were Protestant in their church preference; only 14 per cent were Catholic.

Psychoneurotic Patients

The 34 patients were veterans who had been diagnosed by members of the hospital staff as psychoneurotic. They were selected by a staff psychiatrist as being fully capable of taking the questionnaire. Nearly all were in an open ward of the hospital and could come and go freely; they had been in the hospital less than six months. Forty patients were given the questionnaire; six failed to complete it and their papers were discarded.

All members of the patient group were male, white, and veterans. About half of them were under thirty years of age and nearly all were under forty. Equal proportions of married and single men were included in the group.

In family income the patients were slightly lower than the other groups, but there is some question as to the validity of income data provided by members of this group inasmuch as all were receiving benefits from the government. In relative income they did not differ much from the other groups, but a greater proportion of the patients said they used to earn more than now.

The patient group was lower in education than members of other groups. One-third had less than high-school education, about half had completed high school, and only 12 per cent were college graduates. About two-thirds of them were clerical, operative, service workers, or laborers. Almost one-fourth of the patients failed to answer the question of their father's education, and 12 per cent failed to give their father's occupation, so that these data cannot be meaningfully compared with that of other groups. The distribution of the group in social-class identification and in religious preference was fairly similar to that of the male employee group. Their ratings on their own general health were lower than those of other groups.

---

[1] See Appendix for a more detailed comparison of the sample than summarized here.

EMPLOYEE SAMPLE

The employee sample was selected in such a way as to obtain the best possible representation of the different occupational groups in the hospital with the least possible interference with the activities of the hospital. Sex, income, and occupational description of the major departments was obtained and the questionnaires were allocated within the departments in such a way as to insure representation of the upper and lower occupational groups for each sex.

The sample included 133 white women and 142 white men employees, a total of 275 government employees working in the same general situation. They did not vary as greatly in income and occupation as might be expected in a random sample of the adult population. Only 6 per cent of the employees reported family incomes lower than $40.00 per week; 3 per cent reported incomes of $200.00 per week or more. As government employees, these respondents probably also enjoy more economic and job security than do many of those working in private employment, inasmuch as discharge without adequate cause is not usually possible. Furthermore, employees at the mental hospital are selected with an eye to avoiding applicants who have any history of psychiatric disorder, so as a group the employees were considered to be somewhat more secure than a random sample of the normal adult population.

METHOD OF DISTRIBUTING THE QUESTIONNAIRES

The questionnaires were distributed to the Princeton students during a regular class session. They were handed out with a statement assuring the students that we were not interested in them as individuals, but only as members of a group; instructions were read from this statement for emphasis. The students were asked to seal the questionnaire in the envelopes provided, and to drop them into a locked secret ballot box.

The patients were given the questionnaire in a group situation similar to that of the students. The instructions were read to them, and envelopes were provided for the questionnaires. It was emphasized that they were not being singled out for individual study—that a sample of the staff members was taking part in the same study.

The questionnaires were distributed to the employees under the direction of the chiefs of the major departments in the hospital. The same statement assuring the respondents that we were not interested in them as individuals, and a Princeton University envelope accompanied each ballot. The departmental heads were asked to have the ballots distributed as randomly as possible within the occupational groups selected for study. Excellent cooperation was received from all department heads. For the most part the questionnaires were returned through the department distributing them. Frequently, however, questionnaires given out by one department were returned through another department.

V. THE AREAS OF SECURITY AND INSECURITY

THE concept of personal security accepted as the framework for this investigation recognizes any number of functional areas within which people strive for security. Unique individual patterns of security and insecurity would be expected in addition to the varying levels of security in general. These unique patterns of security are considered to result from individual differences in family background, past experiences, and personal and social purposes. In view of this, differences would be expected in the patterns of security of groups varying primarily in background and status, or "station in life."

It should be recalled that the sample was chosen in such a manner as to obtain the best possible representation of certain major occupational groups within a veterans' mental hospital. Comparisons in terms of sex, age, military service, and other such variables were unwarranted, except when occupation could be held constant. Most comparisons of this nature, therefore, have been postponed until such time as a sample of the national population can be studied.

OCCUPATIONAL CLASSIFICATION OF HOSPITAL EMPLOYEES USED

The manner in which the sample was selected permitted grouping the employees into six occupational groups for a comparison in terms of the total security score, and, also, in terms of the scores on separate parts of the Inventory.

The occupational classification used in grouping the members of the sample was that employed by the Census Department in 1940. Since this classification system is quite thorough, and most of the respondents were fairly specific in describing their work, little difficulty was encountered. In view of the special weighting of certain specific occupations within the major groups, some description of these groups in greater detail may be useful.

The male professional-administrative group included doctors, psychiatrists, psychologists, therapists, social and welfare workers, technical engineers, and departmental administrators.

The female professional-administrative group included trained nurses, therapists, social and welfare workers, dietitians, and departmental administrators. Two-thirds of the group were nurses, and the security scores for the nurses and other women professional and administrative workers were similar, so that nurses were not treated as a distinct group in this analysis.

The male technical, clerical, crafts, and foremen group included technicians, sports directors, and other semi-professional employees, clerical workers, electricians, laundry foremen, painters and carpenters, non-technical engineers, etc. The semi-professional workers were too few in number to treat as a distinct group. As a group they differed but slightly from the clerical workers, craftsmen, and foremen.

The female nonprofessional group included all female employees below the professional level. About two-thirds of the group were administrative assistants, secretaries, stenographers, and clerical workers of various kinds. Also included were laundry workers, seamstresses, waitresses, kitchen help, housekeepers, and hostesses. The distribution of security scores for the clerical workers was very similar to that of other members of this group.

The male attendants and aides were grouped by the Census Department as service employees. Since more than 500 aides are employed by the hospital, a large enough sample was obtained to treat them as a separate group for this analysis.

The male service and operator group included machine operators and machinists, meat cutters, laundry workers, truck drivers, guards and watchmen, barbers, storekeepers, cooks, bakers, and kitchen help.

UNIQUE, INDIVIDUAL PATTERNS OF SECURITY

A comparison of the security scores of the areas of striving covered by the Inventory brought to light a wide variety of individual differences. A number of subjects who ranked most insecure (0) in several areas ranked vary secure (4) in one or two of the areas. Conversely, some professional persons who ranked very secure in nearly all areas were most insecure in one or two.

The free answer questions at the end of the Inventory, some of which were quite revealing, provided an opportunity for unique expression. Projection was clearly in evidence in many of the responses to these questions; for example:

> "What do people need to make them feel secure?"
>
> A male foreman over 50: "Good health." (What else?) "Retirement at the age of 55."
>
> A patient: "Faith in the Lord! Realize that they are temporary custodians of the Lord's footstool and that they are allowed on earth to earn their way to heaven."
>
> "What do you think makes people feel most insecure?"
>
> A male service employee: "Lack of food more than clothes. Being behind in payments."
>
> A patient: "Lack of love and harmony in the home from childhood to manhood."
>
> A woman clerk: "Being a social outcast."

The Areas People Care Most About

Health (being in good physical and mental health) and economic security (being sure of a job or steady income with good living conditions) were the outstanding choices of nearly all groups in response to the question:

"Which three of these do you think are most important? That is, which three do you care most about?"

A. Having a chance to develop some special talent, interest, or hobby.
B. Having a good social standing and the right kind of people as friends.
C. Having a good ability to understand things, good education, and general information.
D. Being sure of a job or steady income with good living conditions.
E. Being satisfied in your relations with members of the opposite sex.
F. Having good looks and appearances, good build, height, etc.
G. Being liked and respected, and feeling you "belong."
H. Having a good religious understanding and faith.
I. Being in good physical and mental health.
J. Being satisfied with your work or job.

Health was chosen as one of the most important areas by the overwhelming majority of the members of all groups. Economic security received second mention by all groups, except the male service employees (1st) and the female professional and administrative employees (4th).

Some variations occurred throughout, but the areas identified in the analysis as Belongingness (G), Capacities (C), Work Satisfaction (J), and Religion (H) were next most frequently mentioned by the employee groups.

There was an interesting relationship between general level of security or occupational status and the selection of Social Status (having a good social standing and the right kind of people as friends). A larger proportion of the insecure groups selected Social Status as an area they cared most about; fewer of them chose it as an area they cared least about. The regression was fairly regular (Table 1).

About one-third of the patients also selected Capacities (having a good ability to understand things, good education, and general information) as an area they cared least about. Not more than 7 per cent of any other group made this choice.

Responses to the questions on the areas cared most about and those cared least about were found to bear a fairly close relationship to the answers to the projective free-answer questions at the end of the interview on what people need for security (Question 38). The number of things mentioned in response to the open question varied greatly. Some subjects failed to give any answer,

TABLE 1
FREQUENCY OF SELECTION OF SOCIAL STATUS AS AN AREA CARED MOST ABOUT OR CARED LEAST ABOUT*

Group	*N*	Cared Most About	*N*	Cared Least About
Male professional-administrative	29	3%	27	63%
Female professional-administrative	64	3	61	64
Female nonprofessional	52	8	53	53
Male attendants	35	14	37	32
Male tech., cler., etc.	28	18	28	50
Male service workers	27	18	29	34
Patients	33	33	31	29
No answer	30		32	
Number of cases	298		298	

* This table is based on responses to Part I of Inventory. The Social Status area was formulated: "Having a good social standing and the right kind of people as friends."

and others mentioned as many as five or six things.

A few subjects also failed to answer the question on the areas they cared most about. Some of these wrote comments such as "All of these areas are equally important," or "I don't care least about any of these." Some checked all of the areas as things they cared most about.

AREAS OF STRIVING TEND TO BE AREAS OF INSECURITY

The preliminary investigations had suggested that the things people say they care most about tend to be the things about which they feel most insecure. The final study yielded further support for this.

In making analysis of the relationship between the areas of striving respondents cared most about and the levels of security within these areas, it was necessary to assume that the groups of questions chosen to measure security feelings within the different areas were of approximately the same discrimination value. If this were not true, any relationship existing might very well be concealed by the inadequacies of the questionnaire.

In constructing the Inventory, four questions were included to measure security within each of the areas, except the religious and economic-security areas.[1] The question objectives for the different areas were similar, even though the actual wordings of the questions varied. In general, for every area one question was concerned each with satisfaction with present status, status relative to associates, feelings of surety, and expectations. Some evidence that the groups of questions representing the different areas have approximately similar discrimination value was the fact that the mean area security scores did not vary greatly either within the most secure groups or within the least secure groups. For the male professional and administrative group, the range in area mean scores was from 2.6 to 3.5; for the neurotic group it was from 1.4 to 2.4.

Rank-order correlations were made

---

[1] No attempt was made to measure religious security. Eight questions were concerned with economic security, and half the mean score was used in making comparisons.

between the areas of striving selected by the different groups and the mean security scores for the areas to test the thesis that subjects tended to be most insecure in the areas they cared most about. The health area was left out in making this analysis because health was consistently ranked highest or second, although the employee groups were all relatively very secure in this area. The religious area was also excluded because the Inventory contained no religious-security score.

Biserial correlations could not be used in making this analysis because the general level of security and also the levels of security within the areas varied directly with economic and occupational status (Ch. VI). Furthermore, the areas of striving cared most about also tended to vary to some extent with occupational status. It was clear that any analysis using biserial correlational techniques would be meaningless unless the sample used were large enough to keep both occupation and general level of security constant. The present sample did not permit this finer control.

This analysis made by occupational group indicated that there was a fairly strong and consistent tendency for the areas of greatest insecurity, as measured by the Inventory, to be the ones cared most about. The rank-order correlations were greater than .65 for most of the groups studied. The exceptions were the male professional and administrative employees (.45), the female professional and administrative employees (.35), and the patients (.39) (Table 2).

Although the correlation between areas cared most about and areas of greatest insecurity for the patient group was only .39, a correlation of .77 maintained between the areas cared least about and the areas of greatest security. Thus the patients tended to be most secure in the areas they cared least about. For other groups, the two approaches yielded essentially the same correlations.

There is evidence that the fairly high rank-order correlations between areas cared most about and areas of insecurity does not reflect a pattern peculiar to the employee group; this relationship was found to be fairly close for Princeton students (.87) and for a smaller group of employees of a social welfare organization, also (.88).

TABLE 2

RANK-ORDER CORRELATIONS (RHO) BETWEEN AREAS OF STRIVING SUBJECTS CARE MOST ABOUT AND AREAS OF GREATEST PERSONAL INSECURITY

	Number of Cases	*Rho*
Male professional-administrative employees	29	.45
Female professional-administrative employees	64	.35
Female nonprofessional employees	52	.76
Male attendants	35	.68
Male technical, clerical, foremen	28	.86
Male service and operatives	27	.86
Princeton students	51	.87
Social welfare organization employees	20	.88
Psychoneurotic patients	33	.39*
No answer to Question No. 1	30	
Number of cases	369	

* Rho between areas patients cared least about and areas of greatest personal security was .77.

PATTERNS OF SECURITY TEND TO VARY WITH STATION IN LIFE

Cantril (3), Horowitz (6), and others have pointed out that the pattern of a person's areas of security and insecurity will tend to be a function of his past experiences, present situation, group relationships, etc. According to this thesis there should be a relationship between occupational status and patterns of security. Groups that are most alike in occupation would be expected to have the most similar patterns of security. An analysis of the present data showed a consistent tendency in this direction.

The mean area security scores of the occupational groups were ranked in the order of security level, with the rank of *1* being given to the area of greatest insecurity as measured by the Inventory.

TABLE 3

RANKING OF MEAN AREA PERSONAL SECURITY SCORES BY OCCUPATIONAL GROUP

Area of security	Male Prof. Adm.	Female Prof. Adm.	Female non-prof.	Male Attend.	Male Tech. Cler. Frmn.	Male Serv. Oper.	Patients	Princeton Students
Economic	1	1.5	1	1	1	1	5	6
Cultural	2	1.5	4	4.5	7	6	7	5
Belongingness	3	3.5	2	2	3	3.5	3	4
Capacities	4.5	5	5	6.5	4	3.5	6.5	1.5
Looks	4.5	8	8	8	9	5	8	9
Social Status	6	6	7	6.5	5	7	2	7
Work Satisfaction	7	7	3	3	2	2	4	1.5
Sex	8	3.5	6	4.5	6	8	6.5	3
Health	9	9	9	9	8	9	1	8

NOTE: The Census Department occupational classification system was used in grouping these hospital employees. See page 20 for a fuller discussion of the occupational groups. In presenting this table, there is one variation from the Census Department classification. The male technical, foremen, clerical, and craftsmen were shifted below the male attendant group. The nature of their work was considered more like that of the service employees than like that of the professional employees.

For the employee groups, economic security consistently received the lowest ranking (insecure), and health was consistently ranked 8th or 9th. Other security areas varied in rank. The patterns of both patients and Princeton students differed consistently from the employee groups and from each other (Table 3).

Rank-order correlations were then made between the mean security scores for the individual areas of striving by occupational group (Table 4). Groups most similar in occupation tend to have patterns of security that are most alike. Conversely, the patterns of security of groups differing widely in occupation tend to be only slightly related. This is also true in the areas they care most about and care least about.

Security in the work-satisfaction area, compared with the security level in other areas, tended to increase with occupational status. At the same time there was a tendency for the security level in the

TABLE 4

RANK-ORDER CORRELATIONS (RHO) BETWEEN MEAN AREA PERSONAL SECURITY SCORES OF DIFFERENT OCCUPATIONAL GROUPS

Group	Male Prof. Adm.	Female Prof. Adm.	Female non-Prof.	Male Attend.	Male Tech. Cler. Frmn.	Male Serv. Oper.	Patients	Princeton Students
Male Prof.-Adm.		.72	.68	.57	.36	.64	−.20	−.01
Female Prof.-Adm.	.72		.73	.74	.45	.35	−.20	−.34
Female Nonprof.	.68	.73		.96	.85	.82	.02	.54
Male Attendants	.57	.74	.96		.82	.70	.06	.49
Male Tech., Cler., etc.	.36	.45	.85	.82		.88	.29	.59
Male Serv.	.64	.35	.82	.70	.88		−.13	.40
Patients	−.20	−.20	.02	.06	.29	−.13		.02
Princeton Students	−.01	.34	.54	.49	.59	.40	.02	

NOTE: See note to Table 3.

cultural-activities area, relative to that in other areas, to decrease with higher occupational status. These trends are considered subjects for further study with a greater and more varied sample.

The data further suggest that insofar as this employee sample is concerned, at least, occupational status and the things associated with it have a greater bearing than sex difference upon patterns of security. The patterns of security of the female nonprofessional employees are more like those of other nonprofessional employees than like those of the female professional and administrative employees. Their group security pattern is most closely similar to that of the male attendants (.96) whose status is most like their own. The group pattern of the female professional and administrative employees, on the other hand, bears about the same relationship to that of the male attendants (.74) as it does to that of the female nonprofessional group (.73).

The patterns of security of the patient group differed greatly from those of all other groups studied, the highest correlation with any group being .29. Thus, the patients were found to be a distinct group both in the low level of security in general as measured by the Inventory, and in the unique pattern of the areas in which they feel most and least secure.

Princeton students differed most from the patients and from the male and female professional and administrative employees. Their group pattern of security bore the greatest resemblance to those of the female nonprofessional, male attendant, and male technical, foremen, and craftsmen groups, but the relationship was not great.

It should not be inferred from these data that the relationships found between occupational status and patterns of security reflect the influence of occupation alone. Occupation, income, education, family background, etc., are all closely related, so that a comparison in terms of occupation provided but one means of getting at what might better be called a person's "station in life." The conclusion here is that an individual's patterns of security and insecurity are closely related to his station in life.

It is important to emphasize that some factors that probably influence patterns of security, such as age, religion, and marital status, could not be held constant in this analysis. On the other hand, the employee sample was considered to be more homogeneous than a sample of the national population would be in terms of age, veteran status, education, income, place of employment, and nature of work.

VI. THE GENERAL LEVEL OF SECURITY

Analysis of the data obtained in the preliminary study indicated the presence of a general security level as well as levels of security within the various areas of striving. At that time questions concerned with subjective economic security and work satisfaction seemed to provide an index to this general security.

In the final study, several methods were used to investigate more fully this relationship between economic security and security in areas of striving. Support was found for the hypothesis that persons who lack economic security tend to feel insecure in the more developmental areas of striving, such as belongingness, sex, personal appearance, cultural activities, etc.

The Problem of Measuring Economic Status

The relationship between economic status and feelings of personal security is a particularly difficult one to investigate, because income is tied up with other important variables, such as abilities and capacities, health, family background, social status, and education. Education and social status, for example, normally parallel economic status in our society; the individual who has wealth yet lacks education and social standing is an exception so far as our culture is concerned.

There is also the problem of determining what an individual's real income is. It is hazardous to compare incomes without taking into account such things as local living costs, the size of family, ages of children, etc. We might agree that a man supporting a wife and a child on an income of $3,000 is better off than his neighbor who is supporting a family of four or five on the same income. It is more difficult, however, to determine objectively how much better off he is, and how his economic needs vary as his family grows up.

A number of subjective factors should be considered in making economic comparisons. There is the problem of present income compared with previous income, and also compared with the income one may expect to have in the future. Furthermore, economic needs vary according to one's position in life and in the community; one's own income relative to that of other members of his membership groups must be considered.

In this study several approaches were used in an attempt to obtain a meaningful answer to the problem of the relationship between economic security and security in other areas of striving. Data were obtained on the income of the immediate family, subjective economic status and work satisfaction, income relative to that of associates, and satisfaction with opportunities for personal development.

Occupational data provided a better measure of personal earnings than did the information on income of the immediate family, because the latter also included the earnings of other members of the family. Occupation is a measure of job status and social status, as well as of economic status, and to a certain degree serves as an index to opportunities for personal development. Most people need a certain amount of family and other support to achieve the higher professional positions. To some extent personal ability is also reflected in occupation.

TABLE 5

COMPARISON OF MEAN PERSONAL SECURITY SCORES OF HOSPITAL EMPLOYEES DIFFERING IN RELATIVE INCOME AND SATISFACTION WITH OPPORTUNITIES

	N	Mean Security Score*	*SD*	*SE* Mean
Dissatisfied with opportunities and—				
Lower relative income	25	21.12	7.58	1.21
About the same or higher	66	22.50	5.52	.68
Satisfied with opportunities and—				
Lower relative income	35	25.38	7.71	1.30
About the same or higher	139	29.01	5.96	.51
No answer	10			
Number of cases	275			

* One point has been deducted from the mean scores of those satisfied with opportunities, because one point was given for a "satisfied" response in computing the total security scores.

question on relative income in making a breakdown of the employee sample, a very decided relationship to level of security was indicated (Table 5). Data on male and female groups were combined in making this comparison since no diference in response was observed.

OCCUPATION[3]

A comparison of occupational groups in term of the mean security scores (Table 6) shows a consistent relationship between level of occupation and level of security, which is true for both male and female groups. Since all respondents were employed in the same hospital, occupation represents both income and job status. The male professional and administrative employees, for example, tend to be the best paid and also hold the most respected positions. Both of these factors need to be considered in interpreting the relationship.

A rise in the mean security score by occupational level is accompanied by a decrease in the variability of the group scores. A number of those in the lower occupational groups had high security

TABLE 6

MEAN PERSONAL SECURITY SCORES

Comparison between Hospital Employees (classified by occupational level), Princeton Students, and Psychoneurotic Patients

Group	*N*	Mean Security Score	*SD*	*SE* Mean	Diff. Sign. from other groups*
Male, prof.-adminis.	30	32.50	4.01	.73	2, 3, 4, 5, 6, 7, 8
Female, prof.-adminis.	67	28.61	7.52	.92	1, (5), 6, 7, 8
Princeton students	51	26.55	5.59	.78	1, 8
Male, technical, clerical, foremen	31	26.29	6.96	1.25	1, 8
Male, attendants and aides	40	25.93	5.67	.90	1, (2), 8
Female, nonprofessional	64	24.63	6.45	.81	1, 2, 8
Male, service and machine operators	32	24.41	7.28	1.29	1, 2, 8
Psychoneurotic patients	34	19.06	10.51	1.80	1, 2, 3, 4, 5, 6, 7

* Parenthesis enclosing the group number, e.g. (2), indicates the .05 level of confidence. Otherwise, the .01 level of confidence is indicated.

---

[3] See page 21 for a detailed description of the occupational classification of hospital employees.

TABLE 7

Mean Personal Security Scores, Excluding Economic-Work-Satisfaction Items
Comparison between Hospital Employees (classified by occupational level),
Princeton Students, and Psychoneuritic Patients

Group	*N*	Mean Score	*SD*	*SE* Mean	Diff. Sign. from other groups
Male, Prof.-adminis.	30	22.93	3.19	.56	2, 3, 4, 5, 6, 7, 8
Female, prof.-adminis.	67	19.82	5.61	.69	1, 7, 8
Princeton students	51	18.51	4.60	.44	1, 8
Male, technical, clerical, foremen	31	19.61	5.02	.90	1, 8
Male, attendants and aides	40	19.10	4.37	.69	1, 8
Female, nonprofessional	54	18.33	4.66	.58	1, 8
Male, service and machine operators	32	17.63	5.31	.94	1, 2, 8
Psychoneurotic patients	34	13.29	7.09	1.22	1, 2, 3, 4, 5, 6, 7

scores; none of the male professional-administrative group had very low security scores.

A possibility was recognized that the differences between groups might primarily reflect differences in responses to the economic- and work-satisfaction questions. If this were true, the differences could be interpreted as reflecting income only. The groups were therefore compared in terms of both their responses to the questions relating to subjective economic security, and those relating to security in other areas of striving.

When this comparison was made (Table 7, Table 8), it was found that the noneconomic questions differentiate between the groups more clearly than do the subjective economic-work-satisfaction questions. The latter questions supplement the rest of the Inventory very well in spreading out the scores. It should be noted that the lower occupational groups do not differ very much from the psychoneurotics on the economic-work-satisfaction questions. This may be attributed to the fact that the psychoneurotic group consists entirely of veterans who are assured of pensions, medical and financial assistance during illness, and aid for limited periods of unemployment, which was not true of the majority of the employees below the professional level.

TABLE 8

Mean Economic-Work-Satisfaction Security Scores
Comparison Between Hospital Employees (classified by occupational level),
Princeton Students, and Psychoneurotic Patients

Group	*N*	Mean Score	*SD*	*SE* Mean	Diff. Sign. from other groups*
Male, prof.-adminis.	30	8.60	1.83	.33	3, 4, 5, 6, 7, 8
Female, prof.-adminis.	67	8.19	2.78	.27	4, 5, 6, 7, 8
Princeton students	51	7.33	2.71	.38	1, (4), 6, 8
Male, technical, clerical, foremen	31	6.06	3.17	.57	1, 2, (3)
Male, attendants and aides	40	6.35	2.36	.37	1, 2, (8)
Female, nonprofessional	64	5.78	2.91	.36	1, 2, 3
Male, service and machine operators	32	6.16	2.56	.45	1, 2
Psychoneurotic patients	34	5.09	2.56	.44	1, 2, 3, (5)

* Parenthesis enclosing the group number, e.g. (4), indicates the .05 level of confidence. Otherwise, the .01 level of confidence is indicated.

EDUCATION

The employee sample used in this study was selected in such a manner as to obtain adequate representation of the various economic and occupational groups within the hospital. In view of this, the educational distribution obtained cannot be considered representative either of the normal adult population or of the employees of the hospital. The upper educational levels are heavily over-represented:

Education	*Male Employees*	*Female Employees*	*All Employees*
No schooling	—	—	—
Grammar School (Gr. 1-8)	18%	7%	13%
Part of High School (Gr. 9, 10, 11) ...	19	9	14
Finished High School (Gr. 12)	21	29	25
Part of College	20	31	25
Finished College ...	22	24	23
No answer	—	—	—
	100%	100%	100%
Number of cases ..	142	133	275

For the men employees no general trend in the relationship between education and the security level was observed; however, a sharp increase in security occurred at the college graduate level (Table 9). There was no such sharp rise in security at the graduate level for the women employees. The trend here was more consistent, with a somewhat greater break occurring between the group without any college education and those who had attended but not finished college. This was believed to reflect the preponderance of nurses among the professional women employees. A college degree is not required of registered nurses, but they are required to have a number of years of specialized training.

The absence of a consistent relationship between education and level of security of the employee group as a whole and for the two sex groups suggests that education, per se, does not have a great bearing on the general level of security, except insofar as additional education brings with it additional income and status. This is a problem for further exploration.

CONCLUSIONS

Studies by Bakke (1;2), Hall (5), Rundquist and Sletto (9), Stauffer and Lazarsfeld (10), Zawadski and Lazarsfeld (11), and others have shown that during periods of unemployment, depression, and economic stress, personalities undergo a severe breakdown. People lose confidence in their abilities, lack feelings of status and belongingness, and sometimes even lose hope.

The present findings differ from these in that they are based on a study of normal, relatively secure, employed people. None of the employees in the present

TABLE 9

MEAN PERSONAL SECURITY SCORES OF HOSPITAL EMPLOYEES
Comparison by Education

Level of Education	*N*	Security Score	*SD*	*SE* Mean
Grammar school—part of high school	74	25.24	7.11	.83
Finished high school	68	25.21	6.46	.78
Part of college	70	26.99	6.80	.81
Finished college	63	30.40	6.21	.78
No answer	—	—	—	—
Total number of cases	275			

study had any history of psychiatric disorder. They did not differ widely in income. Within this relatively homogeneous sample of employed people a consistent relationship was found between economic security (especially, subjectively measured security) and feelings of security in the more developmental areas of striving.

Whether weekly income of the immediate family, subjective measures of economic security and work satisfaction, relative income, relative income and satisfaction with opportunities for development, or occupational status was used as an index to the level of economic security, a positive relationship between economic security and the level of security in other areas of striving was found to exist within the employee sample. The relationship remained when all economic, and work-satisfaction items were removed from the personal security score, to give a more limited measure of security in the other areas of striving.

The possibility existed that the responses to the subjective economic- and work-satisfaction items represented in good part a projection of feelings of security or insecurity in other areas of striving. However, the relationship between economic security in other areas also exists when other measures of security, such as relative income, education, and occupation are used as measures of economic security. Therefore, if projection occurs, it is not wholly responsible for the relationship.

It should be noted that the total security score used in this analysis is a composite of the security scores for nine major areas of striving. No conclusions are drawn from the data in regard to the relationship between economic security and security level within any other single area.

VII. SUMMARY AND CONCLUSIONS

The general purpose of this study was to investigate the nature of personal security to determine what it means psychologically, and to gain some insights about correlates of personal security. This involved formulating a working concept of security from the existing literature and applying it in a concrete situation.

A preliminary concept of security was formulated and was used as the framework in constructing a Personal Security Inventory. The Inventory was then administered to a sample of employees in a veterans' mental hospital and to a group of patients who had been classified by psychologists and psychiatrists as psychoneurotic. It was also administered to a group of students in psychology and to a smaller group of employees of a social welfare agency in New York City.

A person's feelings of security or insecurity involve his own subjective evaluation of his success, satisfaction, and surety or confidence with respect to carrying out his purposes in past and present situations and group relationships; also, his expectations, hopes, fears, or uncertainties about the carrying out of his purposes and aspirations in future situations and group relationships. A questionnaire constructed to measure these subjective feelings was found to discriminate between groups of individuals who were ostensibly secure and those classified by psychiatrists as "insecure" (Chapter III on Validation).

People strive for security. By and large, the things they consider most important, that they care most about, tend to be the things about which they feel most insecure. Conversely, the things they care least about tend to be the areas in which they feel most secure. This relationship, while consistent, was revealed to be somewhat less pronounced for persons of higher professional status, possibly due to a more objective attitude on the part of such persons (Ch. V on Areas).

There are functional areas within which people feel secure or insecure, and unique patterns of security exist within even a fairly homogeneous employee population. In addition to these unique individual patterns of security and insecurity, broader group patterns are present. The patterns of security of very similar occupational groups are more alike than are those of more distantly related occupational groups. Occupation, income, education, etc., were all closely related within the population studied, so that occupational status provided but one measure of what might be called a person's *station in life;* therefore, it is concluded that an individual's patterns of security and insecurity are closely related to his station in life (Ch. V on Areas).

A general level of personal security tends to be present in addition to the levels of security within the individual areas of striving. This seems to occur because the areas of personal striving are interrelated and overlapping, even though some are of greater importance than others to the security of the individual (Ch. VI on General Level of Security).

QUALIFICATIONS

It is important to re-emphasize that these findings are based primarily on a study of three main groups of individuals: employees of a veterans' mental hospital, Princeton University students of psychology, and patients diagnosed by psychiatrists as psychoneurotic. The government employees, who comprised the major sample, did not vary as widely in age, veteran status, income, and occupation as would be expected of the general adult population. It is believed, therefore, that differences found between the occupational groups of this employee sample are relatively less than those existing in more varied populations (Ch. IV on Sample).

On the other hand, special conditions existing within this major sample may be responsible for some of the relationships observed. The findings are considered evidence of probable relationships within the normal adult population which need to be investigated more fully.

THE PERSONAL SECURITY INVENTORY

The Personal Security Inventory developed in the course of this investigation was constructed to measure security feelings in terms of the concept of personal security described above. It was found valid for discriminating between patients diagnosed by psychiatrists as "psychoneurotic" and all occupational groups sampled from among the employee population of the hospital. The patients were found to differ in their scores on this Inventory from other groups, both in terms of their general level of security and in their patterns of security (Ch. VI on General Level and Ch. V on Areas). These validation studies involved groups rather than individuals, and no claims are made beyond suggestive validity of the Inventory in individual diagnosis.

REFERENCES

1. BAKKE, E. W. *Citizens without work.* New Haven: Yale Univer. Press, 1940.
2. BAKKE, E. W. *The unemployed worker.* New Haven: Yale Univer. Press, 1940.
3. CANTRIL, H. Presidential address delivered at annual meeting of Society for the Psychological Study of Social Issues, Sept. 1948.
4. GUNDLACH, R. H. Emotional stability and political opinions as related to age and income. *J. soc. Psychol.*, 1939, 10, 577-590.
5. HALL, O. M. Attitudes and unemployment: a comparison of the opinions and attitudes of employed and unemployed men. *Arch. Psychol., N.Y.*, 1934, No. 165.
6. HOROWITZ, E. L. A dynamic theory of security. *J. soc. Psychol.*, 1939, 10, 421-435.
7. HYMAN, H. The psychology of status. *Arch. Psychol., N.Y.*, 1942, No. 269.
8. KELLEY, T. L. *Scientific method.* Columbus, Ohio: Ohio State Univer. Press, 1932.
9. RUNDQUIST, E. A. & SLETTO, R. F. *Personality in the depression.* Minneapolis: Univer. of Minnesota Press, 1936.
10. STAUFFER, S. A. & LAZARSFELD, P. F. Research memo on the family in the depression. *Bull. Soc. Sci. Research Council*, 1937. No. 29.
11. ZAWADSKI, B., & LAZARSFELD, P. The psychological consequence of unemployment. *J. Sociol.*, 1935, 224-250.

APPENDIX A

TABLE 10

Comparative Description of Samples Studied*

	White Male Employees	White Female Employees	Princeton Psychology Students	Psycho-neurotic Patients
Number of cases	142	133	51	34
Age:				
Under 30 years	35%	49%	100%	53%
Between 30 and 39 years	35	32	—	44
Between 40 and 49 years	15	12	—	3
50 years or over	15	7	—	—
No answer	—	—	—	—
Marital Status:				
Married	68%	32%	12%	44%
Single	25	60	88	41
Divorced	5	6	—	6
No answer	2	2	—	9
Birthplace:				
New York City area (50-mile radius)	47%	48%	39%	44%
Elsewhere in United States	41	48	55	56
Outside United States	12	3	6	—
No answer	—	1	—	—
Military Service:				
Yes	85%	18%	80%	100%
No	15	82	20	—
No answer	—	—	—	—
Father's Education:				
No schooling	12%	2%	—	15%
Grammar school (Grades 1–8)	55	38	8%	32
Part of high school (Grades 9, 10, 11)	9	13	12	12
Finished high school (Grade 12)	11	17	13	9
Part of college	3	8	20	3
Finished college	4	14	47	6
No answer	6	8	—	23
Family Income:				
Between $ 20.00 and $ 29.99 per week	1%	1%	—	6%
Between $ 30.00 and $ 39.99 per week	6	5	4%	21
Between $ 40.00 and $ 59.99 per week	45	22	2	29
Between $ 60.00 and $ 79.99 per week	16	19	8	18
Between $ 80.00 and $ 99.99 per week	14	20	14	3
Between $100.00 and $199.99 per week	13	25	25	15
$200.00 a week or more	1	5	47	3
No answer	4	3	—	5
Average Number of Persons Living on This Income	2.96	2.77	3.6	3.53
Comparison of Present Income with Past Income:				
Used to earn more than now	35%	29%	39%	53%
Used to earn about the same	21	21	29	18
Used to earn less	38	47	27	26
No answer or don't know	6	3	5	3

---

* See text or other tables for description of the samples in terms of education, occupation, and relative income.

TABLE 10—*(continued)*

	White Male Employees	White Female Employees	Princeton Psychology Students	Psycho-neurotic Patients
FATHER'S OCCUPATION:				
Professional workers	8%	15%	27%	3%
Semi-professional workers	1	2	2	—
Farmers and farm managers	6	7	1	—
Proprietors, managers, and officials, except farm	2	9	2	9
Clerical, sales, and kindred workers	8	11	—	2
Craftsmen, foremen, and kindred workers	12	8	1	2
Operatives and kindred workers	25	17	—	32
Service workers	7	7	—	7
Laborers	11	5	2	15
Students	—	—	—	—
Business men—unspecified	6	5	24	—
Executives, bankers, brokers, etc.	—	1	33	—
Small businessmen	9	7	2	18
No answer	5	6	6	12
SOCIAL CLASS:				
Lower class	1%	—	—	—
Working class	53	33%	4%	47%
Middle class	44	61	55	41
Upper class	1	3	39	6
No answer	1	3	2	6
CHURCH PREFERENCE:				
Catholic	43%	45%	14%	41%
Protestant	45	49	67	50
Jewish	9	2	10	6
Other	2	2	9	
No answer	1	2	—	3

APPENDIX B

PERSONAL SECURITY INVENTORY

INSTRUCTION SHEET

This is a Public Opinion Poll

The Office of Public Opinion Research of Princeton University is making a study of what people want out of life—what makes people happy and what makes people unhappy. The Manager has kindly permitted us to interview a large number of persons like you. We need your cooperation.

Please check this public opinion questionnaire in privacy.

Remember: This is a Public Opinion Poll.

DO NOT SIGN YOUR NAME

No one can ever know what you personally answer.

Directions

1. Please read each question through carefully before checking your answer.
2. Answer questions in the order in which they appear.
3. *Answer every question.* This is important. If the answers given do not quite fit your situation, check the one that comes *nearest* to what you want to say.
4. Remember, your own honest and frank answers are the only "right" answers.
5. When you are through, make a double check to be sure that you have not missed any question. But do not make any changes in the answers you have checked.
6. Be careful to follow instructions.
7. When you are through, seal your questionnaire in the envelope.

DO NOT MARK THE ENVELOPE IN ANY WAY

The envelopes will only be opened by the Office of Public Opinion Research.

Thank you for helping us.

Office of Public Opinion Research
Princeton, New Jersey

OFFICE OF PUBLIC OPINION RESEARCH
PRINCETON UNIVERSITY

PART I

Here is a list of things of importance to all of us. But we do not agree about the most important and least important of them.

a. Which three of these do *you* think are *most* important? That is, which three do you care most about?

Make check marks in the first column in front of the *three* things you care most about.

b. Which three do you care *least* about?

Make check marks in the second column in front of the *three* things you care least about.

(There should now be three check marks in each column.)

Care *most* about	Care *least* about	
______	______	A. Having a chance to develop some special talent, interest, or hobby.
______	______	B. Having a good social standing and the right kind of people as friends.
______	______	C. Having a good ability to understand things, good education and general information.

Form A.

Copyright, 1948, Andie L. Knutson, Research Associates, O.P.O.R., Princeton, N.J.

______ ______ D. Being sure of a job or steady income with good living conditions.

______ ______ E. Being satisfied in your relations with members of the opposite sex.

______ ______ F. Having good looks and appearances, good build, height, etc.

______ ______ G. Being liked and respected, and feeling you "belong."

______ ______ H. Having a good religious understanding and faith.

______ ______ I. Being in good physical and mental health.

______ ______ J. Being satisfied with your work or job.

PART II*

At the present time, are you satisfied or dissatisfied with the following things, as far as you personally are concerned?

For example, the first item is:

Are you satisfied or dissatisfied with the amount and kind of food you can afford to buy?

If you are *satisfied*, make a check in the column marked Satisfied.

If you are *not satisfied*, make a check in the column marked Dissatisfied.

If you don't know whether you are satisfied or not, make a check in the column marked Don't know.

(When you are through, there should be one check for each of the eight things listed.)

Satisfied	*Dissatisfied*	*Don't Know*	
1. ______	______	______	The amount and kind of food you can afford to buy.
2. ______	______	______	The spare time you have for the things you want to do.
3. ______	______	______	Your present chances of keeping (or getting) a good job.
4. ______	______	______	The place you can afford to live in.
5. ______	______	______	The clothes you can afford to buy.
6. ______	______	______	The chance you have to get the education you want.
7. ______	______	______	The chances you have of a steady income when you grow old.
8. ______	______	______	The chances you have of a steady income and help if you are sick.

For the following questions, check the answer that applies to you personally. Remember, your own honest and frank answers are the only "right" answers.

9. Taking into account the way your life is and so on, are you satisfied or dissatisfied with the opportunities you have to develop your interests, talents, and abilities the way you would like to?	______Satisfied ______Dissatisfied ______Don't know
10. Does the work you do give you a feeling of self-importance and success, or not?	______Yes ______No ______Don't know
11. Do you have any special interest, talent, or hobby that gives you a feeling of success?	______Yes ______No ______Don't know
12. By and large, are you satisfied or dissatisfied with your health at present?	______Satisfied ______Dissatisfied ______Don't know
13. Do you feel sure of your social acts and manners, or do you feel unsure of yourself socially?	______Feel sure ______Feel unsure ______Don't know
14. Do you think that your looks and appearances have tended to help you, or have tended to stand in the way of your getting ahead?	______Tended to help ______Tended to stand in way ______Don't know
15. Are you satisfied or dissatisfied with the chances you have to get the job you want?	______Satisfied ______Dissatisfied ______Don't know
16. Do you feel sure that people are interested in your ideas and what you are going to do, or do you wonder whether they are interested?	______Sure are interested ______Wonder if interested ______Don't know

---

* Item 22 was not scored in the early form of the inventory described in this report.

17. Do you feel satisfied or dissatisfied in your relations with members of the opposite sex?

______Satisfied
______Dissatisfied
______Don't know

18. Do you feel sure of finishing things you start?

______Yes
______No
______Don't know

19. Are you planning now to do something on your own that you are sure will give you a feeling of success?

______Yes
______No
______Don't know

20. Do you wonder whether people like and respect you?

______Yes
______No
______Don't know

21. Do you feel unsure of yourself because of your looks and appearances?

______Yes
______No
______Don't know

22. On the whole, does life tend to be happy or unhappy for you?

______Happy
______Unhappy
______Don't know

23. Do you feel left out of the groups you go with?

______Yes
______No
______Don't know

24. Are you sure you know what you most want out of life?

______Yes
______No
______Don't know

25. Does the work you do bring out your best talents and abilities, and give you a chance to try out ideas of your own?

______Yes
______No
______Don't know

26. Do the people you go with look up to you for what you know and understand?

______Yes
______No
______Don't know

27. In general, do you feel as healthy as the people you go with, or not?

______Feel as healthy
______Do not feel as healthy
______Don't know

28. Do you think that your looks and appearances are likely to hold you back in the future?

______Yes
______No
______Don't know

29. Have you done anything outside of work that someone you admire has thought worth while?

______Yes
______No
______Don't know

30. Do you feel sure that the people you go with accept you as a social equal, or do you wonder about that?

______Feel sure
______Wonder about it
______Don't know

31. Do you think your social background has tended to help you or to stand in the way of your getting ahead?

______Tended to help
______Tended to stand in way
______Don't know

32. Are you more bothered about sexual matters than the people you go with seem to be?

______Yes
______No
______Don't know

33. As far as your health in general is concerned, does the future look very good, fair, or poor?

______Very good
______Fair
______Poor

34. Do the people you go with give you credit and recognition for what you do, or do they fail to do this?

______Give credit
______Fail to give credit
______Don't know

35. Do the people you work with and for give you credit and recognition for what you do, or do they fail to do this?

______Give credit
______Fail to give credit
______Don't know

36. Do you feel unsure of yourself because of your health?

______Yes
______No
______Don't know

37. Do you feel as successful as the people you go with in the things you do outside of work?

______As successful
______Not as successful
______Don't know

38. When it comes to getting along with members of the opposite sex, do you feel sure of yourself, or unsure of yourself?

______Sure
______Unsure
______Don't know

39. Would you say that the future in this area, that is in getting along with members of the opposite sex, looks very good, fair, or poor?

______Very good
______Fair
______Poor

40. Are you bothered by wanting to do things you do not feel mentally or intellectually able to do?

______Yes
______No
______Don't know

41. Do you feel satisfied or dissatisfied with your present social standing?

______Satisfied
______Dissatisfied
______Don't know

42. Do you think that looks and appearances help to keep you from feeling you "belong?"

______Yes
______No
______Don't know

PART III

1. What do you think people need to feel most secure? ....

....

2. What else? ....

....

....

3. What do you think makes people feel most insecure? ....

....

4. What else? ....

....

....

VOLUME 66
NUMBER 5

WHOLE NO. 337
1952

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Validity of Personality-Trait Ratings Based on Projective Techniques

By

HENRY SAMUELS

Veterans Administration Center
Columbus, Ohio

Based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Michigan

Accepted for publication July 17, 1951

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

THE VALIDITY OF PERSONALITY-TRAIT RATINGS BASED ON PROJECTIVE TECHNIQUES

Chapter I. *Introduction* 1
Chapter II. *Procedures and Methods* 2
Chapter III. *Results* 7
Chapter IV. *Other Factors Related to Validity* 13
Chapter V. *Summary and Conclusions* 19
Bibliography 21

CHAPTER I

INTRODUCTION[1]

CONSIDERABLE importance has been attached to projective techniques by clinical psychologists and professional personnel in allied fields. Such devices have been used in psychiatric diagnosis, academic and industrial selection, educational and vocational guidance, research studies of personality development, and in psychotherapy. One of the outstanding problems with respect to such techniques is their validity in their various applications.

A review of the literature on projection techniques emphasizes the need for determining the extent to which projective techniques agree with independent criteria in the description of personality. We may have different kinds of validities depending upon the function projective techniques are asked to serve.

The purpose of the present investigation is to continue the study of the validity of projective techniques by considering the following questions:

(1) How well are projective clinicians who use the same projective method able to *describe* personality? Are there differences among clinicians in this ability?

(2) Are there differences in the extent to which personality is correctly described which are related to the kind of projective method used?

(3) Are there differences among attributes of personality which make for differences in the degree to which they can be correctly described?

---

[1] I wish to acknowledge my indebtedness to all the people who helped to make this study possible. I am especially grateful to Professor E. Lowell Kelly for his counsel, instruction, and labor, without stint. Such value as may be derived from the study must, for the most part, be attributed to him. I am indebted to Dr. Donald G. Marquis, Dr. George A. Satter, Dr. Max L. Hutt, and Dr. Wm. Clark Trow for their suggestions and criticisms, and to Dr. Donald W. Fiske and Dr. Ernest C. Tupes for their willing help. My wife, Helen, was, as always, inspiring and stimulating, and, in addition, labored long and arduously with the data. Caroline Weichlein and Phyllis E. Moore were most zealous with the secretarial chores.

The opinions expressed herein are the author's and do not necessarily reflect the views of the Veterans Administration.

CHAPTER II

PROCEDURES AND METHODS

The present investigation was done as part of the research project on the Selection of Clinical Psychologists sponsored by the Veterans Administration under a contract with the University of Michigan, under the direction of Professor E. Lowell Kelly (3, 4).

A. THE ASSESSMENT PROGRAM

During the Summer of 1947, 140 students who had been selected by universities for first-year positions in the Veterans Administration clinical psychology training program came to Ann Arbor in "classes" of 24 per week to be assessed. During the assessment week a student completed a battery of paper and pencil objective tests, and had administered to him in group form the Thematic Apperception Test (10 cards) and the project's form of a sentence completion test. During the first day or two the student was administered the Rorschach and the Bender-Gestalt individually and at separate sessions. Each student was interviewed twice (6), filled out a 131-item Biographical Inventory, wrote an autobiography, was put through a variety of situations, and filled out a sociometric questionnaire.

The subjects, on whom the ratings investigated in this study were made, consisted of 128 male college graduates who had been accepted by various universities for graduate training in clinical psychology. The twelve women who were part of the total assessed group have been omitted from this study.

Because they are crucial to the present study, a detailed discussion is given to the rating scales, the projective ratings, and the criterion ratings.

B. THE RATING SCALES

The rating scales cover 42 variables divided into three sections which are given the letter designations A, B, and C.[1] Scale A, consisting of 22 variables, was designed to describe more overt or phenotypical dimensions of personality and was adapted from Cattell's (1) factorial studies, with modifications made on the basis of experience with the scale in earlier pilot assessments. In the present study 5 of the 22 Scale A variables are used. These 5 were selected on the basis of a factorial analysis by Fiske (2). Each of these 5 selected Scale A variables has highest factor loadings for Fiske's five factors which are nearly orthogonal to each other. These 5 selected variables are:

# 4 Depressed–Cheerful
#17 Conscientious–Not Conscientious
#18 Imaginative–Unimaginative
#21 Dependent Minded–Independent Minded
#22 Limited overt emotional expression–Marked overt emotional expression

The definitions of Scale A variables were designed to depict these variables as sharply as possible as surface behavior, i.e., how the individual would appear directly to the observer.

The 9 Scale B variables were designed to provide opportunity for judgment of the more covert, genotypical, dynamic, or interpretive aspects of personality. Despite the known intercorrelations between the Scale B variables (from −.57 to .79) all 9 have been used in this study, since it is in the description of these aspects of personality that projective techniques are commonly thought to be most useful. These 9 Scale B variables are:

---

[1] Complete definitions of all 42 personality traits may be found in a doctoral dissertation by Samuels (5, Appendix A).

- #23 Social Adjustment
- #24 Appropriateness of Emotional Expression
- #25 Characteristic Intensity of Inner Emotional Tension
- #26 Sexual Adjustment
- #27 Motivation for Professional Status
- #28 Motivation for Scientific Understanding of People
- #29 Insight into Others
- #30 Insight into Himself
- #31 Quality of Intellectual Accomplishments

Scale C included 11 variables on which staff members were asked to make predictive judgment. Of these, only number 42, "Over-all Suitability for Clinical Psychology," is used here, since this was the variable toward which the entire assessment program was oriented. Throughout the remainder of the discussion Variable 42 is treated as though it were a Scale B variable for simplification in statistical summary. This assumes that the rating on "Over-all Suitability" represents an inference about a covert, although perhaps extremely complex, dimension of personality and is essentially similar to Scale B variables.

All ratings were made on an 8-point scale with a theoretical distribution of ratings of 3, 7, 15, 25, 25, 15, 7, and 3 per cent for points from 1 to 8 respectively. Raters were instructed to use a reference population of first-year clinical psychology graduate students in American universities.

C. The Projective Ratings

As was stated above, during the early part of the week each student was examined with the Rorschach and the Bender-Gestalt *individually*. Each record was scored, interpreted, summarized, and the applicant rated on the scales described above by the test administrator, without other knowledge about the individual than could be gained from the particular projective technique and the personal contact involved during the testing period. The Thematic Apperception Test and the project form of a sentence-completion test were administered as *group tests*. Each subject was then rated on each variable of Scale A and Scale B by four different staff members, each rater basing his ratings on a different projective technique. A single staff member was concerned with only one projective method per subject and was instructed to keep himself in ignorance of the findings of the other staff members for that subject. After each of the four projective techniques had been independently analyzed and subjects rated on the basis of each, the Rorschach analyst, serving in a new role of "projective integrator," made another set of ratings on each variable of Scale B based on *all* the projective data, with the exception of the actual ratings which had been independently made on the basis of each projective technique.

The number of students seen by a given clinician was not the same for a given technique, nor was the number of students seen by a given clinician the same, necessarily, for two or more techniques. The number of students per rater per technique is shown in Table 1.

A given rater may have analyzed more Rorschachs than another, fewer TAT's than any or all of the others, and some other number of SC's. The differences in N between the two Scales, A and B, are attributable to the fact that cases which were integrated by the "projective integrator" were not rated on Scale A by the "projective integrator" nor by the other projective-technique raters. The projective raters used in this study are fewer

TABLE 1

NUMBER OF STUDENTS RATED BY EACH RATER ON FOUR PROJECTIVE TECHNIQUES

Technique	Rater									
	W	O	L	F	T	G	R	Q	M	P
					Scale A					
Rorschach	11	11	12	10	5	9				
TAT	7	8	9	15	3*	14				
SC	4	8	14	4	13	2**				
BG							19	14	15	13
					Scale B					
Rorschach	18	23	22	22	20	20				
TAT	22	22	14	28	4	25				
SC	16	13	37	4	26	6				
BG							32	29	33	33

* With an *N* of three, *r* is either indeterminate or fluctuates widely.
** With an *N* of two, *r* is .00 or 1.00.

in number than the total number of projectivists on the assessment staff. Only six are used, each of whom rated enough students via Rorschach, Thematic Apperception Test, and sentence completion to make comparisons among them possible. Four other projective raters were concerned only with the Bender-Gestalt. The latter four, coded by letters R, Q, M, and P, were graduate students at the University of Michigan, at the third-year level of training in the VA training program for clinical psychologists, and had been trained in the clinical use of the Bender-Gestalt by the same instructor. The other projectivists are coded W, O, L, F, T, and G. They were initially selected on the basis of professional competence, and the choice of another group of projective-analysis raters similarly selected would probably not have resulted in any better interpretation of the projective tests. The sampling of techniques and the sampling of variables are both regarded as adequate. Raters were given full freedom with respect to scoring, interpretation, and report writing. The only formal imposition placed upon them was the conversion of impressions, deductions, inferences, or conclusions into single numbers on each of the rated variables.

Although students were assigned randomly for projective examination, it was possible that the relatively small number of students seen by a given projectivist using one of the projective devices might have represented a different student population from that seen by another projective rater using the same projective instrument. That is, differences between projectivists' ratings might be found which might be attributable to differences in the students examined and not to differences in the projective techniques. As an additional check on the random assignment of students for projective testing the following test was made. Distributions of the FinP ratings[2] on students on Variable 42, "Over-all Suitability," were made separately for each of the four projective techniques used in this study. For a given technique these ratings for students were distributed for each of the projective raters. In this way it was possible to compare the mean FinP ratings on Variable 42 with respect to the student population seen by each of the projectivists for each of the projective techniques. Epsilon square was computed and no significant differences were found among the mean ratings for students assigned to different projective raters using the same technique. Since the students had been randomly assigned for projective testing and since no differences were found between students for Variable 42, there was no reason to believe that systematic differences would be found between students on other variables, and no other tests were made.

---

[2] Final Pooled Ratings. PI will be used as an abbreviation for the Projective Integrator and the Projective Integration.

D. The Criterion Ratings

In the seven-day period of each student's assessment, three staff members rated the student at frequent intervals. A given staff member would receive certain data on a student and rate, then receive other data and rate again, having filed away his previous ratings, and continue to make independent ratings based on various kinds of data. The first team conference (Preliminary Pooling Conference) produced a single set of ratings which represented a combination of the judgments of the staff team based on these materials: objective and projective tests, credentials, autobiography and Biographical Inventory, and the information obtained by way of the interview (6). At the final pooling conference, all previous material and all previous ratings were made available to the staff team. Also available at this time were the student's self-ratings, ratings of each student by the other three members of the student team, the separate ratings based on projective methods, and the ratings made by a team of staff members who had observed the student in situation tests with no other knowledge about him. A separate final pooled rating was made on each variable of Scale A and of Scale B. These ratings are regarded as the best available measures of each of the 42 personality traits for each subject assessed, and are the criterion measures of this study. They are designated FinP.

The FinP ratings are the most comprehensive and inclusive ratings to come out of the entire week's assessment. These ratings were made by different combinations of three staff members. The staff were initially selected on the basis of professional competence. They had had opportunity to study the student in a wide variety of ways, and had had the benefit of judgments of many other skilled clinicians at various points. Although they are admittedly fallible, there can be little doubt that these FinP raings are as valid criterion measures of these variables as are obtainable at the present time from skilled clinicians using present techniques in an assessment situation.

The FinP ratings are composite criteria consisting of the pooled judgments of three staff members who had available to them the results of the tests and procedures indicated above. A separate FinP rating was made for each variable of Scale A and Scale B. Since staff members were free to arrive at their final judgments as they chose, it would be difficult to say whether FinP ratings emphasized abilities, capacities, past achievement, or personality characteristics. Of interest in this connection is ". . . the fact that assessment staff members tended to be uniformly of the opinion that the interview contributed most to their 'understanding of the case,' followed by either the projective tests or autobiography" (4, p. 404).

Two problems must be considered with respect to the relationship between the ratings based on projective techniques and the FinP ratings. First, how much influence did any or all of the projective techniques have on the criterion measures? Second, are there differences in the criterion ratings when the Projective Integrator is and is not a member of the criterion team?

With respect to the first problem of the influence of projective techniques on criterion measures, a quantitative answer cannot be given. The projective technique protocols and the ratings based on them were available to the final pooling team. This constitutes a theoretical defect in the design of this study which was a sacrifice which had to be made in consideration of other aspects of the over-all project design. It would have been preferable, of course, if the ratings based on each projective technique could have been checked against a criterion based on all available data except that technique. This deficiency with respect to experimental independence would have the effect of spuriously raising the correlations between the ratings based on projective techniques and the criterion. The possible effect of dominance in criterion ratings by any one criterion team member tended to be cancelled by the rotation of staff members from week to week in such a way that for each of the student classes the criterion teams were differently constituted. This does not eliminate the possibility of what might be called group bias with respect to projective methods. For example, many of the staff members, not themselves projectivists, spoke very respectfully of the value of the Rorschach and may have given careful attention to the Rorschach ratings in arriving at the criterion ratings. The possibility exists that another staff, or the same staff assessing at another time, might have another attitude toward projective methods and arrive at a different set of criterion ratings for the same subjects. Since there is small likelihood of another assessment designed to measure the reliability of criterion ratings by the test-retest technique, these obtained criterion measures must be regarded as the only ones available, and therefore, in a practical sense, as universe (not sample) measures.

The problem of the effect of the Projective Integrator as a criterion team member on the criterion team ratings can be evaluated. The Projective Integrator was a member of the criterion team for half the number of poolings. Since the Projective Integrator was also the Rorschach analyst, it might be hypothesized that if he exerted a constant influence on the criterion ratings in the direction of his Rorschach ratings, then the correlations between ratings based on Rorschach and criterion ratings would be higher when the Projector Integrator was a member of the criterion team than they would when he was not. A comparison of correlations between Rorschach ratings and FinP ratings when the Projective Integrator was and was not a member of the criterion team found none of the differences between correlation coefficients under these conditions to be significant at the .05 level, nor was the direction of difference consistent.

CHAPTER III

RESULTS

A. THE PROJECTIVE TECHNIQUES

1. *Validities of ratings based on projective techniques*

TO DETERMINE the extent to which ratings based on projective techniques correlated with FinP ratings, the ratings made by the six raters (four in the case of the Bender-Gestalt) were entered into a single scatter plot for each technique and variable. It was found that something other than "chance" operated to generate statistically significant correlations, since, for Scale A variables, only 1 correlation in 20 was expected to be significant at the .05 level by chance and we find 9 in 20 significant at this level. With respect to Scale B, where 40 correlations were computed, we would expect 2 to reach the .05 level and we find 30 of the 40 correlations are significant at this level. In addition it was noted that only 2 of the 60 coefficients were negative. These two findings, the relatively large number of statistically significant correlations and the high proportion of positive correlations, suggest that the ratings based on projective techniques have some validity. However, these results are, to some extent, spurious since, in addition to the fact that the criterion measures by the six raters were not completely independent, the intercorrelations among Scale B variables were more often positive than negative, and fairly high.

The median correlation coefficients between ratings based on each of four projective techniques and FinP ratings are presented in Table 2. It was noted that correcting the technique-FinP r's for attenuation in the FinP ratings raised the median r only slightly, the greatest increase being from .31 to .35 in the case of the SC on Scale A. That is to say, that even had the FinP ratings been perfectly reliable, the ratings based on projective techniques would have correlated only slightly higher with the FinP ratings.

2. *Differences in the validities of ratings among the projective techniques.*

In order to test the hypothesis that there are no differences in the validities of ratings made on the basis of the four techniques used in this study the following analysis was made. The validity coefficients for each of the techniques for each of the 15 variables were transformed into Fisher's z function and analysis of variance was done for each of the two sets of variables. There was no rater variance here since raters had been combined for each technique. The results of the analysis of variance for the transformed validity coefficients for tech-

TABLE 2

MEDIANS OF CORRELATIONS BETWEEN RATINGS BASED ON EACH OF FOUR PROJECTIVE TECHNIQUES AND FINAL POOLED RATINGS FOR SCALE A AND SCALE B VARIABLES

Scale	Technique							
	Ror.	n	TAT	n	SC	n	BG	n
Scale A	.20	58	.32	56	.31	45	.21	61
Scale B	.29	125	.26	115	.28	102	.14	127
Both Scales	.27		.27		.28		.19	

niques for Scale A showed no significant differences in validity at the .05 level for the techniques or for the variables. The analysis of variance of the Scale B variables showed no significant differences between variables but a difference between techniques significant at the .01 level. Further analysis showed that this difference was attributable to the ratings based on the Bender-Gestalt; this technique differed significantly from each of the other three techniques which did not differ significantly among themselves. Since the Bender-Gestalt raters were four different people from those whose ratings were based on the other three techniques, conclusions from these findings must be made with great care.

3. *Comparison of technique validities on Scales A and B*

When the median validity coefficients,[1] which are given in Table 2, are examined, one is struck by the fact that with the exception of the Rorschach the coefficients are somewhat higher for Scale A, which is not a result that might have been readily hypothesized beforehand. It may also be noted that while many of the median values are statistically significant (.05 level) none of them may reasonably be regarded as predictively useful. Thus, for the techniques studied (with the exception of the Rorschach), we see that projective methods do not lead to more valid inferences about the more presumably covert aspects of personality, as is frequently argued, than they do of the more obvious aspects of personality. On the other hand, and particularly with reference to the TAT and the SC where the student was not seen by the rater (Ror and B-G were administered by the rater) and behaviorial manifestations of the student were inferred from the protocol only, we find that ratings are as valid, by and large, for overt variables as for covert variables.

The possibility that these findings may have been related to the unreliability of the criterion ratings has been shown to be dismissible. Another criticism of these findings may ask if it is fair to require clinicians (in this instance projective clinicians) to express their impressions and judgments in the form of ratings. The core of the argument is that clinicians are accustomed to describing their findings in the form of language (interpretive findings beyond the scores, ratios, percentages, etc. required by the various test instruments) and not in the form of ratings. While this problem must, perhaps necessarily, remain largely a matter for individual judgment, there is considerable merit in the argument that numbers may be more precise than words, and are definitely far more useful in the verification of hypotheses. This position does not deny the possibility that ratings may not accurately represent the judgments about people which clinicians can make from personality test data, nor does it deny the possibility that such judgments may be more accurately communicated in some form other than ratings.

That there are not more differences in validities between the two scales for the projective techniques is a curious finding and difficult to understand. It is not unreasonable to believe that if one has understanding of the *dynamics* of behavior one can infer one step more to the overt behavior of a given individual, but in making the additional inferential step there is additional opportunity for error. It would not have been surprising to find that ratings of overt personality characteristics had less relationship to the criterion ratings than ratings of the covert traits. That they do not—and since in both cases the relationships with the criterion ratings are low—suggests further study to determine, for example, how much correlation there is between the criterion measures and ratings based on observations of the subject in the projective testing situation, and how much with ratings based on projective data without direct observation of the subject.

4. *Correlations between projective technique ratings*

Up to this point the four projective techniques have been considered as generically related. They have been called by a class name and regarded as methods of evaluating personality characteristics. We have found that ratings based on each of the four techniques show but low correlation with the criterion measures. We shall turn to the question, to what degree are personality ratings based on each of the separate techniques in agreement? This was investigated by intercorrelating the ratings, variable by variable, for each of the four techniques. The results are shown in Table 3. While there is a tendency toward positive relationships (101 out of 120) among the independent ratings (no two ratings of a single subject on a given trait were made by the same projective technique rater), there is very little basis for a feeling of confidence that these instruments serve similar functions, i.e., that these instrument-clinician combinations are measur-

TABLE 3

Correlations Between Ratings Based on Each of the Projective Techniques*

	Scale A					
Variable	Ror/TAT	Ror/SC	Ror/BG	TAT/SC	TAT/BG	SC/BG
4	.02	.08	.08	−.05	.11	.14
17	.11	−.01	.11	.09	.13	−.01
18	.11	.21	.10	.23	.06	.09
21	−.20	.22	.06	.05	.10	−.02
22	−.11	−.01	$\underline{.27}$	$\underline{.46}$	.03	.04
	Scale B					
Variable	Ror/TAT	Ror/SC	Ror/BG	TAT/SC	TAT/BG	SC/BG
23	.12	.04	.11	−.07	.04	.04
24	.02	.09	−.04	.00	.06	−.02
25	−.14	.03	−.01	−.05	.02	−.14
26	.08	.00	$\underline{.17}$	.14	.05	$\underline{.19}$
27	.11	.08	−.09	.14	.05	$\underline{.16}$
28	.15	.02	.07	.04	.13	.10
29	.01	.05	−.16	.13	.07	.01
30	.05	.08	−.09	.09	.09	.08
31	.03	.13	−.07	.06	$\underline{.34}$	.15
42	$\underline{.18}$	$\underline{.22}$	.06	−.05	.10	$\underline{.17}$
	Median correlation for					
Scale						
Scale A	.02	.08	.10	.09	.10	.04
Scale B	.06	.06	−.02	.05	.06	.09
Both Scales	.05	.05	.06	.06	.07	.08

* Correlation coefficients which are underlined are significant at the .05 level. For Scale A, N equals 61. For Scale B, N equals 128.

TABLE 4

MEDIANS OF CORRELATIONS BETWEEN PROJECTIVE TECHNIQUE RATINGS AND FINAL* POOLED RATINGS FOR EACH OF SIX PROJECTIVE RATERS

Scale	*Rorschach* Median correlation for												
	W	*n*	O	*n*	L	*n*	F	*n*	T	*n*	G	*n*	All
Scale A	.26	11	.10	11	.05	12	.16	10	.00	5	.43	9	.12
Scale B	.30	18	.21	23	.26	22	.44	22	.48	20	.20	20	.36
Both Scales	.30		.21		.22		.44		.48		.30		.19

Scale	*TAT* Median correlation for												
	W	*n*	O	*n*	L	*n*	F	*n*	T	*n*	G	*n*	All
Scale A	.71	7	.33	8	.37	9	.30	15	—	3	.30	14	.33
Scale B	.26	22	.16	22	.22	14	.44	28	.46	4	.23	25	.28
Both Scales	.28		.21		.26		.40		—		.28		.30

Scale	*SC* Median correlation for												
	W	*n*	O	*n*	L	*n*	F	*n*	T	*n*	G	*n*	All
Scale A	−.19	4	.54	8	.18	14	.82	4	.52	13	—	2	.52
Scale B	.04	16	.36	13	.32	37	.60	4	.45	26	.46	6	.36
Both Scales	.04		.47		.30		.74		.47		—		.37

*Correlation coefficients which are underlined are significant at the .05 level.

ing the same thing. Although the proportion of positive correlations is large, only 9 correlations out of 120 are significant at the .05 level, where 6 correlations are expected to reach this level by chance.

B. THE PROJECTIVE RATERS

1. *Relative validities of ratings for individual raters*

Beyond the techniques per se we are interested in the relative validities of the individual raters by technique and by variable. In Table 4 may be found the median validity coefficients for all traits in Scale A and Scale B respectively, and for both Scales together. These are the median correlations for each of the six raters for ratings based on the Rorschach, TAT, and SC, respectively. Table 5 contains the same information for the Bender-Gestalt.

The results obtained (of which only the medians are presented here) suggest that two of the six Rorschach analysts

TABLE 5

MEDIANS OF CORRELATIONS BETWEEN BENDER-GESTALT RATINGS AND FINAL POOLED RATINGS FOR EACH OF FOUR BENDER-GESTALT RATERS

Scale	Rater								
	R	*n*	Q	*n*	M	*n*	P	*n*	All
Scale A	.23	19	.00	14	.27	15	.30	13	.25
Scale B	.14	32	.10	29	.22	33	.12	33	.15
Both Scales	.14		.10		.26		.20		.16

were able to make ratings which correlated with the criterion ratings beyond chance expectancy at the .05 level. For these two raters, F and T, the median coefficients for Scale B are, respectively, .44 and .48. These are the only median coefficients which are statistically significant for the six raters on both scales. The number of statistically significant individual (not median) correlations is, considering both scales, greater than would be expected in chance occurrence for any rater, since at the .05 level only one in twenty coefficients is expected to reach this level by chance. The median value, however, is regarded as a better representation of the validity of a rater, in that it allows for the fact of intercorrelation between the traits within each Scale.

While the evaluation of a correlation coefficient in terms of what may be called its social significance is quite subjective, there is little reason for enthusiasm toward the efficiency with which these six Rorschach analysts were able to predict the criterion ratings. Although the validity coefficients for Rorschach raters may seem somewhat discouraging, it is emphasized that the ratings of Rorschach projectivists F and T appear to be reasonably valid. This finding is congruent with the situation which exists in the area of projective testing in which, to a very considerable extent, adeptness in the use of projective methods is acquired through instruction by masters. In the practical situation of assessment, however, four of the six Rorschach raters, initially selected on the basis of clinical competence, did not make as valid ratings as the other two. The high proportion of positive correlations suggests that, under assessment conditions, the Rorschach method provides a basis for some validity of rating.

These findings and observations also apply, in general, to the data obtained for raters using the TAT, SC, and BG.

C. Relative Contributions of Raters, Techniques, and Traits to Validities of Ratings

In order to determine the relative contributions of each of the three sources of variance investigated in this study to the validities of ratings, the following analysis was carried out. The validity coefficients were regarded as scores, and, after being transformed into their respective z functions, were treated by the method of analysis of variance. Analyses were done separately for Scales A and B, and additional analyses for the Bender-Gestalt.

For Scale A traits the results indicated that significant differences in validity may be attributed both to techniques and to raters (.05 level) and that the interaction of techniques and raters makes for differences in validity (.01 level). That is to say, that for Scale A traits, the traits themselves do not contribute significantly to differences in validity. However, the different techniques make for differences in the validity of ratings and the different raters make for differences in validity of ratings. We find differences in validity dependent upon who rates, and upon what technique is being used, and differences in validity for the combinations of raters and techniques.

A similar analysis of Scale B traits showed significant differences in the validity of ratings for technique-rater combinations. Further analysis, however, showed this difference to be largely the contribution of rater differences in validity of rating. As with the Scale A traits, the traits themselves are not a source of differences in validity of ratings. While the techniques did contribute to differences in rating validity for Scale A traits, they are not sources of difference for the Scale B traits. For the presumably covert aspects of personality it appears that differences in validity depend chiefly upon the rater.

Separate analyses of variance for each of the two Scales for the Bender-Gestalt showed that neither variance due to raters nor variance due to traits was significantly greater than variance due to chance. Since only the one projective method was involved in these analyses, there was no technique variance.

It seemed reasonable to expect that some personality traits would be relatively more validly rated than others by one or all of the projective techniques or by one or all of the projective raters. The absence of statistically significant differences in validities of ratings attributable to differences among traits is necessarily qualified in that to the extent to which there is correlation among traits, it is more difficult to find existing differences. This is not a particularly cogent objection in the case of the Scale A variables for which the intercorrelations were low. The Scale B traits do have, in some instance, sizable correlation between them, resulting in an underestimation of the significance of the between-variable variance.

CHAPTER IV

OTHER FACTORS RELATED TO VALIDITY

In this chapter are presented data with respect to factors, other than those originally considered in the scope of this study, which may have had an effect upon the validity of ratings based on projective techniques. Consideration will be given to the possibility of bias in rating; to differences in dispersion (confidence) of ratings; to some additional factors concerning raters; and to the possible effects of unreliability in the predictor and criterion ratings.

A. BIAS IN RATING

One factor which may be related to the validity of ratings based on projective techniques is the frame of reference in which ratings were made. By frame of reference is meant an attitude of optimism represented by a tendency to make ratings in the direction of the "laudatory" or socially desirable end of the rating scale. It is possible that if ratings on a given technique tend, consistently, to be either more or less "laudatory" than the Final Pooled Ratings, that this attitude might be related to the validity of such ratings. Similarly, it is possible that such differences in attitude might be related to differences between the relative validities of ratings among the techniques. In order to determine whether such "frame of reference" differences did obtain, the following was done. First, in order to compare the ratings by each of the techniques with the FinP ratings, the significance of the difference between the mean ratings was calculated. This was done by the usual method of obtaining the quotient of the difference between means divided by the standard error of the difference. In each instance the FinP means were subtracted from the technique means and the arithmetic sign recorded in order that the direction of difference be known. This procedure was followed for the four techniques for the fifteen traits studies. A comparison of the FinP means with those of the Projective Integrator was necessarily limited to the ten Scale B traits. Second, the mean ratings by each of the techniques on the fifteen traits were compared with the mean ratings by each of the other projective techniques, again by calculating the significance of each of the differences.

The results of the comparison of mean ratings for techniques and FinP ratings are presented in Table 6 in the form of an abbreviated table of signs. Either a plus or a minus sign is entered in the table where a statistically significant (.05 level) difference between means occurs. The plus sign is used to indicate that the technique mean was in the "laudatory" direction, while the minus sign indicates that the FinP mean was in the "laudatory" direction. An examination of Table 6 reveals that the number of statistically significant differences found exceeds the number which might have been expected if only chance were involved but these differences are not consistently in the same direction. In the case of the Projective Integrator the obtained significant differences are always in the direction of being less "laudatory," and for the Rorschach one may note a trend in the same direction.[1] These data may be of interest insofar as

---

[1] Soskin (5) analyzed ratings based only on student behavior in a series of standard situations and found that Situationists "see the sample of subjects are being characterized predominately by the condemnatory pole. . . ."

TABLE 6

TABLE OF SIGNS SHOWING STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN MEAN TECHNIQUE RATINGS AND MEAN FINAL POOLED RATINGS*

Variable	Ror-FinP	TAT-FinP	Sc-FinP	BG-FinP	PI-FinP
4					#
17	+		+		#
18					#
21					#
22		+			#
23					−
24	−				−
25				+	−
26			+		
27			−		
28					
29	−		−		−
30	−				−
31		+	+		
42					−

* In this table the "+" sign indicates that there is a statistically significant difference (.05 level) with the technique mean in the "laudatory" direction. The reverse is true for the "−" sign.

# The Projective Integrator did not rate Scale A variables.

they characterize the PI and the Rorschach as tending to look on the dark side of personality, but probably have little, if any, relation to validity. When the differences shown in Table 6 were compared with the validity coefficients which are statistically significant for the techniques and for the Projective Integrator, it was apparent that statistically significant validity coefficients were obtained at least as frequently when there was not a difference between means as when there was a mean difference. It may also be noted that statistically significant validity coefficients are not consistently associated with differences either toward or away from the "laudatory" end of the rating scale.

TABLE 7

TABLE OF STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN MEAN RATINGS BASED ON EACH OF THE PROJECTIVE TECHNIQUES*

Variable	Ror-TAT	Ror-SC	Ror-BG	TAT-SC	TAT-BG	SC-BG
4						
17						+
18						
21						
22	−	−				
23						−
24			−	+		
25			−			
26		−				+
27		+	−	−		−
28				+		
29	−			+		−
30	−		−	+		
31			−			
42						−

* In this table the "+" sign indicates that there is a statistically significant difference (.05 level) with the first of the two techniques at the head of the column in the "laudatory" direction. The reverse is true for the "−" sign.

In Table 7 are presented the signs of statistically significant differences between the mean ratings among the techniques. In this table the minus sign indicates that the first of the two technique abbreviations at the head of the column has a mean rating in a less "laudatory" direction. Thus, the minus sign for Trait #22 in the first column signifies that there is a significant difference between the mean ratings of the Rorschach and TAT on this trait, and that the difference is in the direction of the Rorschach being less "laudatory." The apparent tendency for ratings based on the Bender-Gestalt to be more often in the "laudatory" direction than ratings based on the Rorschach or Sentence Completion may suggest a relationship to differences in validity, since both the Rorschach and Sentence Completion yielded more and higher statistically significant validity coefficients than the Bender-Gestalt. This argument is weakened, however, by the observation that the TAT too yielded more and higher statistically significant validity coefficients than the Bender-Gestalt without there being any differences between the means for these two techniques. Again, one might characterize the Rorschach raters as tending to view people darkly, while the Bender-Gestalt raters tend in the direction of rating personality in a relatively benign fashion.

TABLE 8

TABLE OF STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN STANDARD DEVIATIONS OF TECHNIQUE RATINGS AND FINAL POOLED RATINGS*

Variable	Ror-FinP	TAT-FinP	SC-FinP	BG-FinP	PI-FinP
4					#
17					#
18	+				#
21					#
22					#
23				+	
24	+				
25	+	+	+	+	
26					
27				+	
28		−			
29		+			
30					
31				+	
42			−		

* In this table the "+" sign indicates that there is a statistically significant difference (.05 level) with the technique ratings having the larger standard deviation. The reverse is true for the "−" sign.

# The Projective Integrator did not rate Scale A variables.

B. DISPERSION (CONFIDENCE) OF RATINGS

Another factor which needs to be evaluated, since it might be related to validity of ratings, is the spread of scores or ratings. When there is more spread, the likelihood of obtaining higher correlation coefficients is increased. If one is willing to assume that where a rater lacks confidence in his rating he will tend to rate closer to the mean than he will when he does feel confident, then we can use a measure of spread as an index of confidence in evaluating confidence in rating as a factor in validity.

Table 8 is a table of signs showing where significant differences in spread (standard deviation) occur. A plus sign

TABLE 9

TABLE OF STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN STANDARD DEVIATIONS OF RATINGS ON EACH OF THE PROJECTIVE TECHNIQUES*

Variable	Ror-TAT	Ror-SC	Ror-BG	TAT-SC	TAT-BG	SC-BG
4						
17						
18		+				
21						
22						
23						
24		+			−	−
25						
26						
27			−			
28						
29					−	
30				−		−
31		+				
42	+	+	+	+	−	−

* In this table the "+" sign indicates that there is a statistically significant difference (.05 level) with the first of the two techniques at the head of the column having the larger standard deviation The reverse is true for the "−" sign.

indicates that the spread for the technique ratings is significantly greater than that for the FinP ratings, and vice versa for the minus sign. These data suggest that dispersion of ratings has little or no relation to validity. The ratings based on the Bender-Gestalt, which are the least valid of the four sets of ratings, most often show significantly greater spread than the FinP ratings whereas the Projective Integration ratings, which are more valid for more traits than the ratings based on the Bender-Gestalt, are in no instance significantly different in spread from the FinP ratings. It is interesting to note that ratings based on each of the four projective techniques (but not the Projective Integration) have significantly larger standard deviations on Trait #25, "Characteristic Intensity of Inner Emotional Tension," than the FinP. If this represents a feeling of confidence, it appears not to be justified consistently since the TAT and BG ratings on this trait did not correlate significantly with the criterion ratings.

A comparison of the techniques with each other in terms of willingness to spread ratings was carried out. The results are presented in Table 9. The data suggest that ratings based on the Rorschach tend to be made with more spread than those based on the Sentence Completion and TAT. Similarly, the Bender-Gestalt tends to inspire dispersion of ratings at least equal to the Rorschach. This may be a function of the fact that both the Ror and BG were administered by the rater; i.e., face to face contact may be related to dispersion of ratings (confidence). This would be consistent with the staff's opinion that the interview contributed most to their understanding of the case.

C. ADDITIONAL FACTORS CONCERNING RATERS

It has been shown, (Chapter III, B, 2), that different raters contribute significantly to differences in validities of ratings. One possible explanation of this finding, that "good" raters may have been assigned to one technique and "poor" raters to another, is not too ten-able since all the projective raters used the same techniques (except for the Bender-Gestalt) and, in addition, cases were randomly assigned to raters. On the other hand the number of cases seen by each projective rater varied and the obtained difference in validities for raters may have been related to difference in size of samples. As a way of evaluating whether there are differences in validity associated with the numbers of cases seen by "good" and "poor" raters, the rank-order positions of raters was used as an index of ability to rate. If we consider that those raters in the upper half of the ranking are "good" raters and those in the lower half are "poor" raters, then for Scale B, we find the following. For the Rorschach the "good" raters W, F, and T, saw 65 of a total of 125 cases; for the TAT the "good" raters, W, F, and T, saw 54 of the 115 cases; for the SC the "good" raters, F, T, and G, saw 36 of 102 cases. Also, the size of the median validity coefficient for the SC compares favorably with those of the Rorschach and TAT. We may conclude that the size of sample for different raters is not significantly related to validity of rating.

Another possibility to be considered is the effect of a given projective technique on a rater. There may be something in the nature of a given technique which effects a change in the clinician who uses it. The data in Table 6 suggest that the Rorschach is more often associated with a tendency to see people in a less benign light than are the other techniques. When the Rorschach rater in his role of Projective Integrator makes a new set of ratings, incorporating the results of the other projective techniques with his own Rorschach results, this tendency to see people in a less benign light is increased.

D. Further Evidence on Inter-Judge and Inter-Technique Agreement

As was reported in Chapter III, A, 1, this study found that unreliability of criterion ratings has very little effect on the validity of ratings based on projective techniques. There is no way of knowing, from the present data, to what extent the low validities found are a function of the unreliability of raters. The question may be asked in two ways: to what degree do judges agree when they use the same technique, to what degree does a judge agree with himself using different techniques? In order to answer these questions another study was carried out and some preliminary results were available at the time of writing. In this new study four projectivists made ratings based on each of four projective technique records from each of a total of 20 subjects. The number of subjects and raters was limited by practical considerations. On each of the 16 days of the experiment a rater made ratings based on five projective-technique records from 5 different subjects. For example, on the first day Rater A would have worked with the Rorschach records of Subjects 1 and 16, the TAT of Subject 10, the SC of Subject 12, and the B-G of Subject 4. This same "block" of records would be given to Rater B on the eighth day, to Rater C on the fifth day, and to Rater D on the sixteenth day. In this way each of the 16 blocks of five projective-technique records would be worked with by each of the four raters, so that at the end of the 16 days each rater would have rated all the projective records on all 20 subjects. Tables of random numbers were used to assign the projective technique records to the 16 "blocks," after which the blocks were treated as units and distributed by random numbers to the raters.

TABLE 10

MEDIAN INTERCORRELATION COEFFICIENTS FOR EACH OF FOUR RATERS WHEN RATINGS BASED ON FOUR TECHNIQUES ARE INTERCORRELATED*

Variable	Rater			
	A	B	C	D
4	.32	.12	.19	.06
18	.08	.31	.23	.17
23	−.06	.20	−.03	.02
25	.04	−.15	.06	.18
31	.02	.32	$\underline{.51}$	.20
42	.05	.26	.16	.14

* Correlation coefficients which are underlined are significant at the .05 level.

TABLE 11

MEDIAN INTERCORRELATION COEFFICIENTS FOR EACH OF FOUR TECHNIQUES WHEN RATINGS BY EACH RATER ON THAT TECHNIQUE ARE INTERCORRELATED*

Variable	Technique			
	Ror	TAT	SC	BG
4	.28	$\underline{.49}$	.34	.18
18	$\underline{.52}$	.22	$\underline{.78}$	$\underline{.44}$
23	.35	.20	.36	.20
25	.22	.40	.17	.20
31	$\underline{.58}$	.16	$\underline{.51}$	.34
42	$\underline{.56}$	.34	.34	$\underline{.48}$

* Correlation coefficients which are underlined are significant at the .05 level.

The ratings were then examined for consistency of rating for a rater over the four techniques, and for the consistency of ratings on a given technique by the four raters, in both cases for 20 subjects.

In Table 10 are shown the results for the consistency of the individual raters, i.e., the intra-rater consistency over all four techniques. The data are presented in the form of median values for the 6 intercorrelations of the four techniques. The single correlation which appears to be significant (Rater C, Trait #31) may be meaningless in that at the .05 level 1 correlation in 20 may be expected to reach this level of significance.

In Table 11 are shown the inter-rater (intra-technique) agreements. These again are median correlations for the 6 intercorrelations of the ratings of four raters for the same technique. It appears, from the data in these two tables, that a projective technique tends to generate more consistency in rating behavior than the use of a single projective rater. Stated another way, it would appear that different raters will tend to rate a subject in the same way if they use the same projective technique. A given rater will tend to rate the same subject in a nonconsistent manner when he uses different techniques.

CHAPTER V

SUMMARY AND CONCLUSIONS

A. OBJECTIVES

THE principal objectives of this study were:

1. To determine the relative validities of ratings based on four projective techniques when these were used to describe personality and to predict "over-all suitability" as a clinical psychologist;

2. To determine whether there were differences in the relative validities of ratings based on projective techniques which were ascribable to the rater;

3. To determine whether there were differences in the relative validities of ratings based on projective techniques which were ascribable to the technique;

4. To determine whether there were differences in the relative validities of ratings based on projective techniques which were attributable to the personality trait being rated.

Other related problems were also investigated.

B. METHODS AND PROCEDURES

During an intensive assessment program, a total of 128 male first-year graduate students in clinical psychology were rated by ten staff members on 15 personality variables after analysis of protocols from each of four projective techniques, and again after all the projective material (except the actual ratings) from the four techniques had been examined by a "projective integrator." These ratings were correlated with another set of ratings which were arrived at in a "pooling" conference of three staff members who had studied each subject intensively for a week. Since staff members had *all* the information it was possible to obtain about a subject during an assessment week, the ratings which the three staff members agreed upon in their final conference were taken as the best-rated descriptions of a subject's personality and used as criterion measures. These criterion measures were "contaminated" in that they included the protocols and ratings of the predictors, the effect of which would be spuriously to raise the correlations between ratings based on projective techniques and criterion measures. Under these conditions the correlations between ratings based on projective techniques and the final staff ratings were regarded as validity indices. By comparing the correlation coefficients which were obtained in this manner, the objectives of this study could be achieved.

C. SUMMARY OF FINDINGS

1. *The projective techniques.*

   a. Ratings based on the projective techniques correlated significantly with the criterion measures more frequently than was expected by chance. These correlations were preponderantly positive, but were, by usual standards, low.

   b. Correcting the criterion measures for attenuation raised the median correlations between ratings based on projective techniques and the criterion measures only slightly. The greatest increase was from .31 to .35.

   c. Ratings based on the Rorschach and the Bender-Gestalt (made by the raters who had administered these techniques) tended to be made with greater confidence, i.e., showed greater spread, than ratings based on the Thematic Apperception Test and the Sentence Completion (which were rated "blind").

d. Ratings based on the Rorschach tended to be made away from the "laudatory" end of the ratings scale when compared with ratings based on the other projective techniques and with the criterion ratings.

e. There appeared to be little or no relation between validity of ratings and bias of ratings toward either the "laudatory" or "derogatory" ends of the rating scale.

f. There appeared to be little or no relation between validity of ratings and spread of ratings.

g. Preliminary findings suggested that there was more correlation between ratings made by different raters using the same projective technique, than between ratings made by a single rater using different projective techniques.

2. *The projective raters.*

a. Differences in the validity of ratings based on projective techniques were found which could be attributed to the individuals making the ratings.

b. Significant differences in the validity of ratings obtained not only between raters, but for the interaction between rater and technique, for both scales and for the Rorschach, the Thematic Apperception Test, and the Sentence Completion; interaction could not be tested for the Bender-Gestalt.

c. Although some raters were "better" than others, in no instance did the median validity coefficients of statistical significance exceed $r = .48$.

3. *The Traits.*

No significant differences were found in the validity of ratings which were attributable to the differences in personality traits rated.

D. Conclusions

The conclusions which are drawn from the findings of this study are:

1. Projective techniques, used in the assessment of personality characteristics, measure very little in common.
2. There are significant individual differences in the ability to make valid ratings of personality traits from projective techniques, which appear to be independent of the technique used.
3. The dispersion of ratings made by clinicians on personality traits on the basis of projective techniques does not appear to be related to the validity of ratings.
4. The assessment of the degree to which a subject possesses socially desirable personality traits is, in part, a function of the projective technique which is used in making the assessment.
5. The value of projective techniques as instruments for the assessment of specified personality traits in a group of normal superior adults is apparently limited by low validities.

BIBLIOGRAPHY

1. Cattell, R. B. *Description and measurement of personality.* Yonkers-on-Hudson: World Book Co., 1946.
2. Fiske, D. W. Consistency of factorial structures of personality ratings from different sources. *J. abnorm. soc. Psychol.* 1949, **44**, 329-344.
3. Kelly, E. L. Research on the selection of clinical psychologists. *J. clin. Psychol.* 1947, **3**, 39-42.
4. Kelly, E. L. & Fiske, D. W. The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist,* 1950, **5**, 395-406.
5. Samuels, H. An analysis of some factors affecting ratings of personality traits based on projective techniques. Ph.D. Thesis, Univer. Mich., 1950.
6. Soskin, W. F. A study of personality ratings based upon brief observations of behavior in standard situations. *Microfilm Abstr.* (Ph.D. Thesis) University Microfilms, Ann Arbor, Mich. Publ. No. 1208.
7. Tupes, E. C. An evaluation of personality trait ratings obtained by unstructured assessment interviews. *Psychol. Monogr.* 1950, **64**, No. 11 (Whole No. 317).

VOLUME 66
NUMBER 6

WHOLE NO. 338
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Rorschach Estimates of Personality Attributes in the Michigan Assessment Project

By

WOODROW WILBERT MORRIS
Psychopathic Hospital and College of Medicine
State University of Iowa

Based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of Michigan

Accepted for publication, August 13, 1951

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

ACKNOWLEDGEMENTS

To Professor Harold Bechtholdt for his help in the statistical phases of this study. To Professor Max Hutt, chairman of my committee.

To Professor E. Lowell Kelly, Assessment Project Director, and Professor Donald W. Fiske, Assistant Project Director, for their kindness in making these data available to me. To Dr. Wilbur R. Miller, Director of the Psychopathic Hospital, State University of Iowa, who aided materially in the completion of this work.

This work could not have been completed had it not been for the constant encouragement and patient assistance of my wife; to her I am most grateful.

W. W. M.

Iowa City, Iowa
June 8, 1951

RORSCHACH ESTIMATES OF PERSONALITY ATTRIBUTES IN THE MICHIGAN ASSESSMENT PROJECT

Chapter I. *Introduction* 1
Chapter II. *Procedures and Methods* 1
Chapter III. *Results* 10
Chapter IV. *Summary, Conclusions, and Suggestions for Further Research* ... 20
Appendix A. *The Rating Scales* 23
References 27

CHAPTER I

INTRODUCTION

THE present research stems directly from the University of Michigan Assessment of Clinical Psychologists Project which is being conducted under contract with the Veterans Administration. This is a large-scale, joint research program designed to determine experimentally the necessary qualifications of a clinical psychologist. In its very design it offers numerous subsidiary opportunities to test the validity of the techniques employed in the project. One of these techniques is the Rorschach method, which is the one studied in this investigation.

The purposes of this research are: (*a*) to demonstrate that by using data obtained by means of the Rorschach method, valid estimates of personality variables may be obtained; and (*b*) to compare clinical estimates with those based upon specific Rorschach scores or patterns of scores reputed to be descriptive of the variable in question.

CHAPTER II

PROCEDURES AND METHODS

A. THE UNIVERSITY OF MICHIGAN PROJECT IN THE SELECTION OF CLINICAL PSYCHOLOGISTS[1]

THE present description will be limited to the Ann Arbor program carried out during the summer of 1947, since the data which serve as the basis of the study come from that phase of the total research project. As described in the project's first *Progress Report and Preliminary Fndings* (6), the professional staff in 1947 consisted of thirty members, including two VA psychiatrists and several VA psychologists. Other staff members were psychologists drawn from five different universities and several other institutions. This resulted in considerable diversity of background, training, and viewpoint.

The subjects were student VA trainees,

---

[1] This was a five-year study conducted under contract with the Veterans Administration dating from September 1946, under the direction of Prof. E. Lowell Kelly, Project Director.

already accepted for training by one of the cooperating universities. They came to Ann Arbor in six successive classes of 24 each at nine-day intervals, each class remaining for seven days. Each class was divided into six teams of four students each. Each student team was studied intensively by a staff team of three members.

B. The Assessment Project Rating Scales

A rating scale which would permit of clinical judgments on a variety of personality attributes was first tried in the 1946 preliminary program and revised for the 1947 program. The ratings on this revised rating scale form the basis for the assessment research (see Appendix A).

This rating scale has three parts: *Scale A* was intended to cover primarily phenotypic variables, *descriptive* of the subject as seen by the staff at the time of assessment. Including 22 variables, it was constructed with the help of Professor Raymond Cattell on the basis of his several factorial studies of personality ratings (2).

Scale B was intended to provide judgments of more genotypic variables, i.e., *evaluative* judgments of broader underlying variables, again as seen by the staff at the time of assessment. It includes 10 variables.

Scale C includes a list of 11 criterion variables on which staff members were asked to make predictive judgments relative to the future performance of each candidate (five years after assessment).

All ratings were made on an 8-point scale, with a theoretical distribution of ratings as follows:

8 Upper 3% of a population of first-year graduate students.

7 Upper 7% of a population of first-year graduate students.

6 Upper 15% of a population of first-year graduate students.

5 Upper 25% of a population of first-year graduate students.

4 Lower 25% of a population of first-year graduate students.

3 Lower 15% of a population of first-year graduate students.

2 Lower 7% of a population of first-year graduate students.

1 Lower 3% of a population of first-year graduate students.

For Scale C, the reference population was VA trainees five years after assessment. Staff members were instructed to give anyone who they believed would fail in the training program a rating of 1, 2, or 3 on Item 42 ("over-all suitability for clinical psychology").

C. The Candidate's Schedule

Upon arriving in Ann Arbor and reporting to the sorority house being used for housing and the actual assessment research, the assembled candidates were given a welcoming speech by the Project Director, who indicated the general nature of the program. They were told in a general way what was expected of them and assured that strict professional confidence would be maintained for all personal material gathered by the project.

That same morning the group was given a battery of objective tests. At lunch, as at all meals during the first five days, each team sat at a specified table, along with any two staff members who would not be evaluating that team. In the afternoon of the first day, the class continued taking group tests. Before the end of three days, the following tests had been taken: the Thematic Apperception Test (10 cards), the Project Sentence Completion Test, Miller Analogies, Thurstone's Primary Mental Abilities, the Allport-Vernon Study of Values, Strong Vocational Interest Blank, Guilford-Martin Battery (Inventory of Factors STDCR, GAMIN, and Personnel Inventory I), and the Minnesota Multiphasic Personality Inventory. During the third day or, sometimes, the next day, each subject took the Rorschach and the Bender-Gestalt Test, each at separate sessions.

The first afternoon, each student had an "Initial Interview" lasting one hour. The interviewer, unknown to the subject, had already studied the student's "credentials file," i.e., materials he had submitted with his application to the university. These included the transcript of his college grades, his Civil Service Form 57, and letters of recommendation submitted in his behalf. This interview was searching but not deep lest anxiety be aroused in the subject.

During the evening of the first day, each subject filled out a Biographical Inventory, which contained over a hundred multiple-choice items about background, personal history, experiences, and preferences. He also received an outline for an autobiography he was to write within the next two days.

On the third day, more objective tests were administered including the Kuder Preference Record and the ACE Culture Test. The Strong was taken twice more: once, he filled it out as he thought it would be answered by "women in general" and finally as he thought it would be answered by "men in general." These results were not made available to the assessment staff.

Late that afternoon, each subject met his "intensive interviewer," the staff member whom he would get to know best and the person whose job it was to get to know the assigned student best. This interview was thorough and probing. By the nature of the interviewer's questions, it was obvious to each student that the interviewer had carefully examined his credentials, autobiography, Biographical Inventory, test scores, and the other materials on which the student had spent three days.

The fourth day was free for relaxation and sightseeing. The fifth day was spent in a variety of situational procedures which were observed by staff members. These situations were of three kinds: (1) a group, or team situation, in which from four to eight students participated jointly, (2) dual situations, and (3) individual situations, in which each student participated alone, in the presence of staff members only.

On the morning of the last day, the students gathered in the group testing room for their last written work, which was the completion of a sociometric questionnaire concerning his reactions to his classmates, and a rating of himself and his three teammates.

Following this, he had a final individual appointment with his interviewer. Usually the student went to this interview expecting to learn the staff's opinion of him, but came away with relatively little information, such as test scores, etc.

D. THE STAFF SCHEDULE

Much of the staff time allotments can be judged from the foregoing description of student activities. Only the more important aspects of staff activities not clearly indicated above will be discussed in this section.

During the earlier portion of the week the intensive interviewer prepared for his intensive interview with the student assigned to him by studying, in order, (1) the credentials file, (2) the objective test profiles and answer sheets, (3) the autobiography and biographical inventory, (4) the projective test protocols and their interpretations. After each of these four sets of materials he made ratings on Scales B and C. Thus, each of the four sets of ratings was based on a cumulative collection of materials. Each rating was filed before making his next rating.

The projective tests were handled as follows: each student took both the Rorschach and the Bender-Gestalt Tests individually. Each was interpreted and summarized by the administrator of the test. The Thematic Apperception Test and the Sentence Completion Tests were administered as group tests and were analyzed by various staff members. *No staff member analyzed more than one projective test for any one subject.* In addition to the qualitative interpretation, the projective analyst made ratings on the variables of Scale B and Scale C on the basis of each projective protocol. Thus, four sets of ratings on Scale B and Scale C each, were based on individual projective techniques alone. In addition, a "projective integrator" made ratings based on all of the projective data collected.

The interviewer usually spent about four hours preparing for the intensive interview. After the interview, which lasted about two hours, ratings on Scale B and Scale C were made once again.

At this point, each team of three staff members held a conference to discuss their four students. These preliminary conferences had two purposes: the first was to discuss the case and, by subjectively pooling the three individual and independent ratings on each variable, to arrive at a pooled or team rating on each variable; the second was to provide for a full exchange of data and opinion about each case—with both the initial and the intensive interviewers reporting on their interviews.

On the following day, each staff team observed its four students in the situational procedures. Separate ratings were not made after each situation; staff team ratings were made from notes the next day on the four students observed. The preliminary pooling conference consisted in each staff member's making a set of ratings based on his total impression of the student up to and including the situation tests. The interviewers than held Final Interviews with the students whom they had first seen in Intensive Interviews. The substance of this interview was reported to the other two team members. The staff members then examined the tabulations of the sociometric questionnaire and studied the teammates' ratings of each other and the student self-ratings. They also read the character sketches to determine how each student appeared in the eyes of his teammates and to estimate the insight of the author of each sketch. Also, for the first time, they were permitted to examine all the ratings which had previously been made on each subject. Upon the basis of all of these data a new rating, the Final Individual Rating on each variable, was made. These last were taken by each staff team member to the Final Team Conference at which the staff team made its best judgments about each case *in the light of the entire week's assessment.* This comprehensive report considered many aspects of the person: his emotional life, his intellectual functioning, his relations with others, his motivation for clinical psychology, etc. From a research point of view, its most important function was to record qualifications about certain ratings: e.g., the staff might rate a certain case with the understanding that he would soon be undergoing psychotherapy.

The ratings coming out of this conference for each student were designated the *Final Pooled Ratings.* Since it is the most comprehensive and inclusive set of ratings which was made during the program and since it involved the pooled judgments of skilled clinicians who had been afforded ample opportunity to "get to know each subject as God knows him," it is proposed to use those ratings as criterion ratings in the present investigation.

E. SUBJECTS, DATA, AND STATISTICAL DESIGN

In the present study, the subjects consist of a group of 120 male students assessed during the 1947 summer program in Ann Arbor. All were student VA trainees who had already been accepted for training by one of the cooperating universities.

The data to be analyzed are drawn from three sources: (1) the final pooled ratings (criterion ratings) on the ten variables comprising Scale B, and on Variable 42 of Scale C ("over-all suitability for clinical psychology"); (2) the ratings on the same variables made by the Rorschach examiners, based upon information they obtained from the Rorschach records; and (3) Rorschach scores and patterns obtained directly from the Rorschach records.

Only Scale B variables are being used because Rorschach ratings were not always made on Scale A during the assessment program and because it is felt that most of the items on Scale C are too specifically oriented toward professional skills and, hence, remote from the usual type of data reportedly gleaned from Rorschach results. On the other hand, it is rather universally held that the Rorschach method taps more covert, genotypic personality attributes and this is the definition of the Scale B variables.

Six Rorschach examiners administered approximately four records each during each week of assessment. Each scored the records, following whatever system was familiar to him, most using the scoring symbols designated by Klopfer (3), although some used Beck's categories (1), while one used a system combining features of both of these. For the sake of uniformity, all scores were studied by the writer and converted to those of Klopfer insofar as information available in the records permitted.

It is proposed to correlate the ratings made by the Rorschach examiners (hereinafter designated "Judges") with the final pooled ratings (hereinafter designated "Criterion") variable-by-variable, in order to determine the extent to which judgments based on Rorschach information will approximate the criterion judgments. Since six judges are involved in making the Rorschach ratings, the degree to which the criterion is estimated by the individual judges will be studied by means of running correlations between the ratings of each judge and the criterion ratings. This will result in an array of six correlations for each variable and each of such arrays will be tested by Rider's technique (5) for determining the homogeneity of a distribution of correlation coefficients.

In half of the criterion teams the Rorschach examiner was a functioning team member. Since this introduces the possibility of contamination in the criterion ratings, the data will be split into contaminated and uncontaminated groups and separate correlations computed for each group. Under ordinary circumstances of contamination it might be expected that the contaminated data would show higher correlations than the uncontaminated.

The study of individual Rorschach scores and patterns descriptive of the several variables involved will be discussed individually.

Social Adjustment: The scores and patterns to be used in predicting this variable (Variable 23b) are listed below. The rationale behind the selection of these particular Rorschach factors lies in the interpretative values assigned to them in the literature.

1. Normatively, the *number of responses* usually given to the ten ink-blot designs by healthy, normal adults ranges from 30 to 40. In varying degrees of social maladjustment, it has been observed that the number of responses frequently is less than this, often ranging downward from 25 responses. Miale and Harrower-Erickson (4), for example, report a heterogeneous group of neurotics as giving an average of only 13 responses, only three cases producing more than 25 responses. Their normal control group averaged 36 responses by comparison. Decreased productivity is often interpreted as reflecting one who is ill at ease and lacking in drive and spontaneity. In the present investigation, therefore, the tendency to give less than 25 responses will be taken as a partial indication of social maladjustment.

2. *Human movement responses* are ordinarily interpreted as indicating the subject's degree of inner maturity and stability. Normally two or more of such responses are expected; less than this number suggests a lack of these characteristics. Miale's neurotics, for instance, averaged only 1.1 *M* while the normal controls averaged 5.5 and all had produced two or more of this type of response. Less than two *M* responses will be regarded as a second partial sign of maladjustment.

3. *Animal movement responses,* designated as *FM*, are interpreted as more primitive, infantile inner drives in contrast to the more mature human movement responses. Normally it is expected that in healthy adults, *M* will exceed *FM,* usually by two to one. In Miale's neurotic group, 67 per cent of the records showed *FM* to be greater than *M;* the normals produced more *M* than *FM* in 70 per cent of the records. Indications of such an excess of immaturity in the inner life in the records of the assessment group will be taken, therefore, to be related to social maladjustment.

4. *Color responses* in general in Rorschach interpretation are reputedly related to emotional responsiveness to affectively toned stimuli from the subject's environment. Form responses are interpreted as indicating intellectual control. Hence, responses determined by a combination of form and color, with form elements predominating (*FC*), are reflective of controlled emotionality, hence, emotional adjustment. Two or more of such responses are normally expected. Neurotics, according to Miale's study, produce an average of only 0.6 *FC* responses while her normal controls averaged 3.1; stated differently, 81 per cent of the neurotics gave no more than one *FC* and an equal per cent of the normals gave more than one. In fact, 70 per cent of the neurotics produced no *FC* responses at all. For this reason, less than two *FC* responses will be regarded as indicative of a lack of emotional adjustment and more than two as suggestive of good emotional adjustment and good rapport with others.

5. Another pattern, related to the previous one, but more indicative of emotional imbalance, excitability, and egocentricity is given by the extent to which responses in which intellectual control is subordinated to emotionality (*CF*) or in which intellectual control is completely lacking (*C*) exceed the more controlled *FC* responses. Such *imbalance* and *egocentricity* is posited as indicating social maladjustment.

6. Normally 30 to 50 per cent of Rorschach responses involve the *perception of animal figures.* A per cent of responses in excess of 50 per cent is regarded as mental stereotypy. In 58 per cent of Miale's neurotic cases animal responses constituted over half of the responses; only 25 per cent of the normals showed this sign. Thus, over 50 per cent animal responses in the present records is interpreted as indicating a lack of spontaneity associated with social maladjustment.

7. Also related to spontaneity in Rorschach interpretation is the *per cent of responses determined by form alone.* Too high a per cent of form responses is interpreted as a restriction of spontaneity, with an attempt to substitute intellectual control for genuine adjustment. Normally, 30 to 50 per cent of the total number of responses is expected to be determined by pure form. Fifty-

one per cent of Miale's neurotics showed this sign compared with only 20 per cent among the normals. Following this, subjects who produced over 50 per cent form responses are regarded as having one more sign of social maladjustment.

8. *Color shock* has long been regarded as indicative of emotional disturbances and inhibitions in response to emotionally toned stimuli emanating from the environment. Color shock was defined in the present study as follows:

a. A significant delay, over one average deviation, in time of responding to color cards, especially II and VIII.

b. Exclamations or comments by the subject felt to be indicative of disturbance in reaction to these cards.

c. Decline in the quality of responses, deterioration of mode of approach sequence, or failure to produce commonly perceived responses on the color cards.

d. Rejection of color cards with no scorable response. Using similar criteria for color shock, Miale found the reaction to be present in 98 per cent of the neurotic records and in only 20 per cent of the normal records.

9. *Shading shock,* referred to by Klopfer as "sexual" shock, are reactions similar to those described above to the two highly shaded cards (IV and VI) and occasionally to the diffusely shaded card VII. Such a reaction is interpreted as indicative of considerable anxiety, often related to sexual conflicts. This sign was present in 81 per cent of Miale's neurotic records and in only 20 per cent of the normal ones. The presence of either color shock or shading shock is regarded as indicative of poor social adjustment.

10. Finally, the *complete rejection of any card* with no scorable response is rarely observed among normal, healthy adults. Of Miale's psychoneurotics, 47 per cent rejected one or more cards, the range was from one to seven of the ten cards. None of her normal controls showed this type of reaction. Rejections in the present study will be regarded as another sign of maladjustment.

These "signs" of social maladjustment will be treated in two different ways. The absence of signs will be correlated with the criterion ratings to determine the extent of concomitant variations between them by means of the usual Pearson product-moment method. Second, since Miale and Harrower-Erickson (4) suggest that the presence of five or more signs strongly indicates social maladjustment, a biserial correlation coefficient will be computed, the dichotomous variables being designated as records having ing less than five signs and records having five or more signs of maladjustment.

Appropriateness of Emotional Expression: This variable is closely related to certain of the patterns enumerated above, specifically the response categories designated *FC, CF,* and *C.* The first is ordinarily interpreted as indicative of appropriate emotional responsiveness while the latter two suggest progressively more disorganized emotionality. The correlations between these individual factors will be determined and then combined to determine the multiple correlation coefficient obtaining among them and the criterion ratings. In addition, patterns 4, 5, 7, and 8 noted above under Social Adjustment are felt to be related to appropriateness of emotional expression. A biserial coefficient of correlation will be run between these patterns and the criterion; one dichotomous variable will be the presence of less than two of these patterns, the other two or more. The former is expected to be related to more appropriate emotionality, the latter with failures to adapt emotional responses to the needs of the situation.

Characteristic Intensity of Inner Emotional Tension: Classically, in Rorschach interpretation, human movement responses (*M*), and more recently animal movement responses (*FM*) and the perception of inanimate objects in motion (*m*) have been regarded as associated with the inner life of the individual. The former is deemed representative of a more mature, stable stage of development in the intrapsychic life while the latter two represent progressively a more immature, inhibited, and tension-laden state. Hence, these three factors in the records will be correlated individually with the criterion ratings on this variable (Variable 25) and then combined to determine the extent of multiple correlation among them and the criterion. Since a low rating indicates minimum tensions and a high rating suggests strongly repressed drives resulting in inner turmoil, a negative correlation coefficient should be expected for the correlation with *M*, a positive one with *m*, with *FM* taking a middle position.

Another pattern will be studied here involving the conception noted above that *M* should exceed *FM* normally. To study this pattern, the writer devised an index expression which results in values ranging from zero to two as follows: $1 - (M\text{-}FM)/(M+FM)$. Scores closer to zero would represent the preponderance of *M* in the record while scores approaching two reflect the excess of *FM* over *M*. Negative correlation between these values and the criterion ratings would be expected.

These same factors will be treated in still another way. A biserial correlation coefficient will be computed with $M \overset{=}{<} FM$ and $M > FM$ as the dichotomous variables.

Since inner emotional tensions are plausibly related to social adjustment (Variable 23b), it is proposed that the ten patterns used there be employed again here. In this instance it would be expected that the more signs of maladjustment, the higher should be the criterion rating on inner emotional tension.

Sexual Adjustment: Two Rorschach scores are often used to estimate sexual adjustment developmentally. Animal movement responses (FM) are frequently interpreted as indicative of more basic, instinctual, sexual drives which in adults may be interpreted as a fixation at an early level of psychosexual development. Another way of gauging this factor is in comparison with the more mature inner drives as reflected by human movement responses (*M*). Thus, the number of *FM* responses in the records and the above-described index showing the relative weights of *M* or *FM* will each be correlated with the criterion ratings on Variable 26. It would be expected that each should yield a negative correlation with sexual adjustment if the interpretative assumptions indicated above are to be retained.

Sexual shock reactions, defined under the section on Social Adjustment, are ordinarily interpreted as indicative of anxiety over sexual conflicts and, hence, of sexual maladjustment. In order to assess the value of this pattern a point biserial coefficient of correlation will be run against the criterion ratings, the dichotomous variables being presence and absence of sexual shock in the records.

Motivation for Professional Status: There are no factors in the Rorschach which are presumed to be related to such a variable as this one. However, "drive for status rewards" is suggestive of Rorschach interpretations involving the tendency to over-emphasize responses to the whole ink blot (*W*), frequently referred to as "ambition." Whole responses have also been compared with human movement responses and norms are given in the literature which indicate that the healthy balance between these two factors is approximately 2:1. When *W* exceeds expectations, "ambition" over and above basic creative capacities is inferred. Some Rorschach examiners also feel that this same personality complex is indicated by a large number of responses to the whole Rorschach experiment. Therefore, it is proposed that four correlations be run against the criterion ratings on Variable 27: (1) number of responses utilizing the whole blot (*W*), (2) per cent of responses utilizing the whole blot (%*W*), (3) scores given by the formula (*W*-2*M*), and (4) total number of responses given (R). In each instance positive correlation coefficients would be expected.

Motivation for Scientific Understanding of People: Again, this represents a variable which has no clear counterpart in Rorschach scores or patterns. It may be assumed, however, that scores interpreted as reflecting interest in and empathy for others, and thinking which is in keeping with that of one's contemporaries might be plausibly related to this variable. Content categories involving the perception of human beings are an example of the former type while popular responses, i.e., those commonly seen by other persons, are an example of the latter. Thus, three correlations will be run against the criterion ratings on Variable 28: (1) number of responses in which the content was human beings (*H*), (2) per cent of responses in which the content was human beings (%*H*), and (3) number of popular responses (*P*). In each instance positive correlations might be expected.

Insight into Others: There are no particular Rorschach factors specifically related to this variable insofar as the literature is concerned; this is equally true of the next variable, *Insight into Himself*. However, reference has been made from time to time to a more restricted kind of insightfulness, i.e., insight into anxieties. The presence of this personality characteristic is inferred from responses determined in part by form in combination with either diffuse shading and/or surface texture in the ink blots. Diffuse shading (*K*) is interpreted as anxiety in the inner life, hence, the combination of form and diffuse shading (*FK*) represents anxiety over which the individual is able to exert intellectual control and, hence, anxiety into which he has insight. Surface texture, especially in relation to dysphoric concept, is interpreted as anxiety in relation to environmental occurrences; the combination of form and texture (*Fc*) suggests situational anxieties under intellectual control, or anxieties into which the person has insight. Reasoning similarly, form and color responses (*FC*) may be interpreted as controlled, or insightful, emotional responsiveness to environmental stimuli. These patterns will be used in two ways: (1) the sum of responses in which form elements were primary and combined with textural and diffuse shading determinants (*Fc* + *FK*), and (2) the sum of responses in which form and color were combined determinants, form predominating (*FC*). These will be employed in both Variables 29 and 30 since they seem so closely related by definition and since preliminary study reveals the correlation between them to be .80.

Quality of Intellectual Accomplishments: Intellectual quality in Rorschach work is ordinarily inferred from the quality of the subject's responses. Quality is defined as the degree to which a given response is apposite to reality and is designated as *F* +. In general, the higher the per cent of *F* + responses, the better the quality of the subject's intellectual work. In addition to this factor, originality of responses of good quality (*O* +) is interpreted as high-level intellectual ability. Originality is ordinarily defined as a response given approximately once in a hundred records. A negatively oriented factor related to this variable is the percentage of animal forms (%*A*) seen. This score is interpreted as reflecting mental stereotypy when the percentage exceeds half of the total number of responses given. Thus, four correlations will be run against the criterion ratings on Variable 31: (1) per cent of *F*+ responses, (2) the raw number of original responses given (O), (3) the number of original responses of good quality given (O+), and (4) the percentage of animal responses (%*A*). In the first three instances, positive correlations should result; in the latter, the correlation should be negative.

Over-all Suitability for Clinical Psychology: This variable is obviously neither a specific personality characteristic nor is it one which has ever been related to Rorschach results. However, one might posit the hypothesis that over-all suitability is logically related to over-all social adjustment. Working on this hypothesis, it is proposed that the ten patterns of adjustment used in studying Variable 23b be used here. It will be remembered that the absence of these signs was felt to be related to adjustment, hence, one should expect a positive coefficient of correlation in this case if the hypothesis is to be retained.

CHAPTER III

RESULTS

IN THIS chapter we compare the criterion ratings on each variable with (*a*) the ratings made by Rorschach examiners, and (*b*) the various Rorschach scores and patterns.

A. VARIABLE 23b: SOCIAL ADJUSTMENT

Social adjustment was defined for the raters as follows: *"How well does he adjust to varied interpersonal situations? (Includes sexual adjustment only as it affects social adjustment in general.)"* On this variable, the higher the rating, the more it indicated a person who "actively considers the feelings of others; readily gains acceptance in interpersonal relationships; maintains a friendly and likeable manner." The lower the rating, the more it indicated one who "acts without consideration for the feelings of others; often rejected by others; often appears aloof, hostile, or irritable."

The correlations between the criterion and ratings by the Rorschach "judges" are shown in the first column of Table 1.

On this variable the Rorschach judges were able to predict the criterion ratings to the extent indicated by an r of .35. This can be interpreted as indicating a positive relationship and one which is significantly greater than zero at better than the 1 per cent level of confidence.

As will be noted in Table 1, there is apparently a wide spread of r's among the six judges (from .26 to .63), and one is tempted to conclude that Judges A and B were able to estimate the criterion ratings better than the other four judges. However, Rider's method of testing the homogeneity of such a distribution of r's yields a chi-square value of 4.27, which with five degrees of freedom gives a p of .50, indicating that the array of r's is probably homogeneous. The inference of a difference between one judge and the others might, therefore, be erroneous.

In Table 2, while the contaminated r appears to be greater, the null hypothesis that there is probably no difference between these correlation coefficients may be retained ($p = .16$).

The ten Rorschach scores and patterns described in the preceding chapter were studied in an effort to determine to what extent scores alone might be used to predict the criterion ratings on "social adjustment." The absence of these signs

TABLE 1

CORRELATIONS BETWEEN RORSCHACH RATINGS AND CRITERION RATINGS

Judges	N	Obtained correlation coefficients by variables*										r required at—	
		23b	24	25	26	27	28	29	30	31	C42	.05 level	.01 level
A	23	*.61*	.37	*.49*	*.45*	.02	*.43*	.28	*.55*	*.44*	.25	.41	.52
B	16	*.63*	*.81*	.11	.37	*.60*	.47	.41	.09	*.65*	*.67*	.50	.62
C	19	.32	*.56*	.31	.17	.26	.44	.24	.10	.27	*.65*	.46	.58
D	21	.28	.28	−.20	.23	.38	.38	−.03	*.48*	.15	.20	.43	.55
E	19	.26	.44	.43	.38	*.64*	.24	.24	.30	*.47*	*.46*	.46	.58
F	19	.27	.18	.08	.15	−.13	.32	.27	.27	.25	.23	.46	.58
Total	120**	*.35*	*.43*	*.24*	*.26*	*.32*	*.38*	*.23*	*.27*	*.33*	*.35*	.18	.24

---

* Correlations which are statistically significant at the .05 level of confidence or better are printed in *italics*.

** N does not total 120 since one judge rated only three subjects and is excluded from the analysis of individual judges but is included in the total correlational analysis.

TABLE 2

EFFECT OF CONTAMINATION ON THE CORRELATIONS BETWEEN RORSCHACH RATINGS AND CRITERION RATINGS

Variable	Contaminated *r*	Uncontaminated *r*	Diff.	Critical ratio	*p*
23b	.443	.278	.165	1.02	.16
24	.500	.560	−.060	0.48	.31
25	.348	.163	.185	1.06	.15
26	.118	.403	−.285	1.64	.05*
27	.469	.193	.276	1.66	.05
28	.541	.497	.044	0.33	.37
29	.174	.280	.106	0.59	.28
30	.119	.400	.281	1.62	.05*
31	.258	.367	−.109	0.62	.27
42	.280	.391	−.111	0.66	.25

* While this difference is significant at the .05 level of confidence, note that it is in the opposite direction from that hypothesized.

was correlated with the criterion ratings in two different ways; the results are presented in Table 3. The Pearson *r*, computed from a scattergram, yielded a coefficient of .47; this is somewhat better than was obtained for the typical judge's ratings ($r = .35$).

The use of these ten factors was studied in another way, since in the literature it has been suggested that the presence of five or more of them is a critical number of differentiating adjustment from maladjustment. The computation of biserial correlation produced an $r_{bis} = .55$, which is significantly greater than zero at better than the 1 per cent level of confidence.

B. VARIABLE 24: APPROPRIATENESS OF EMOTIONAL EXPRESSION

This variable is defined as: "*How appropriate are his emotional responses to the situation?*" The lower the ratings the more it indicates an individual who "fails to adapt his emotional responses to the needs of the situation; shows disorganized or overly constricted emotional responses." The higher the rating, the more it indicates one who "shows emotional responses of a quality and intensity befitting the situation; reacts spontaneously but appropriately; shows well-integrated and flexible patterns of emotional behavior."

The correlations between the Rorschach judges' ratings and those of the criterion are shown in column two of Table 1. The results here are similar to those obtained on Variable 23b. The *r* is slightly higher and more in keeping with the lower levels of validity obtained on many psychological tests; again, the total *r* is positive and significantly greater than zero at better than the 1 per cent level of confidence.

The chi-square test of homogeneity of the distribution of *r*'s among the six judges yielded a chi-square value of 8.07 and a *p* of .15. This also may be interpreted as indicating a homogeneous array of *r* values.

The investigation with respect to contamination yielded results which may be interpreted as indicating no significant contamination effect.

Various Rorschach scores and patterns were studied in relation to this variable. First, the three factors *FC*, *CF*, and *C* responses were employed. Using the following indices,

1 The criterion ratings to be predicted,
2 *FC* responses,

3 *CF* responses,
4 *C* responses, the following zero order *r*'s were obtained:

$$r_{12} = .21 \quad r_{23} = .12 \quad r_{34} = .20$$
$$r_{13} = .28 \quad r_{24} = .01$$
$$r_{14} = .02$$
$$M_1 = 4.20 \quad M_2 = 2.97 \quad M_3 = 2.66 \quad M_4 = 0.23$$
$$\sigma_1 = 1.15 \quad \sigma_2 = 2.38 \quad \sigma_3 = 2.66 \quad \sigma_4 = 0.63$$

None of the three zero order *r*'s approaches the size of the correlation between ratings by the judges and the criterion ($r = .43$), although those involving *FC* and *CF* are significantly greater than zero. The multiple correlation coefficient involving the above correlation matrix yielded $R_{1.234} = .33$.

Second, the four "signs": (1) *FC* responses less than two, (2) %*F* responses greater than 50 per cent, (3) presence of

TABLE 3

COMPARISON OF ESTIMATIONS OF CRITERION RATINGS BY CLINICIANS AND FROM RORSCHACH PATTERNS

Variable	Estimate of Criterion by Judges (Table 1)	Estimate of Criterion by Rorschach Indices and Patterns		
				r_{bis}
23b. Social Adjustment	*.35**		*.47*[a]	*.55*[b]
24. Appropriateness of Emotional Expression	*.43*	*FC*:	*.21*	
		CF:	*.28*	*.48*[d]
			.33[c]	
25. Intensity of Inner Emotional Tension	*.24*	*M*:	*−.27*	.09[g]
		FM:	−.04	
		m:	.11	
			.35[e]	
			−.06[f]	
26. Sexual Adjustment	*.26*		*.23*[a]	
		M:	−.09	*.37*[h]
		FM:	−.02	
27. Motivation for Professional Status	*.32*	*M*:*FM*:	.05	
		W:	−.04	
		%*W*:	−.07	
		R:	−.12	
28. Motivation for Scientific Understanding of People	*.38*	*W*−2*M*:	−.17	
		H:	.16	
		%*H*:	.07	
29. Insight into Others	*.23*	P:	−.16	
		Fc+*FK*:	.16	
		FC:	.17	
30. Insight into Himself	*.27*	*Fc*+*FK*:	.08	
		FC:	.17	
31. Quality of Intellectual Accomplishments	*.33*	%*F*+:	.04	
		%*A*:	*−.18*	
		O:	.07	
		O+:	*.19*	
C 42 Over-all Suitability for Clinical Psychology (five years hence)	*.35*		*.52*[a]	

* Italicized coefficients are significantly greater than zero at the 5 per cent level or better ($r_{.05} = .18$, $r_{.01} = .24$).

[a] Ten Rorschach patterns of adjustment (see page 5).

[b] Five or more of these ten patterns and less than five used to dichotomize the group (page 6).

[c] Multiple correlation involving the criterion, *FC*, *CF*, and *C* (page 11).

[d] Four Rorschach patterns of emotionality; group dichotomized on the basis of one or less and two or more patterns (page 6).

[e] Multiple correlation involving the criterion, *M*, *FM*, and *m* (page 13).

[f] Index written: $1-(M-FM)/(M+FM)$ (page 7).

[g] Group dichotomized on the basis of $M<FM$ and $M>FM$ (page 7).

[h] Point biserial correlation; group dichotomized on the presence or absence of sexual shock reactions (page 7).

color shock reactions, and (4) *FC* less than the sum of *CF* plus *C* patterns were studied in relation to the criterion ratings. A biserial r was computed, the dichotomous variables being the presence of less than two "signs" on the one hand and two or more on the other. This yielded a biserial r of .48, which is significantly greater than zero at better than the 1 per cent level of confidence.

C. VARIABLE 25: CHARACTERISTIC INTENSITY OF INNER EMOTIONAL TENSION

This variable is defined as *"How intense is his inner emotional life as* inferred *from all available clues?"* Here a low rating suggests an individual whose "inner emotional life is characterized by a minimum of persistent internal tensions." A high rating indicates one who "has strongly repressed emotional drives resulting in inner turmoil; great inner conflict and strong pent-up emotions."

Column three of Table 1 presents the correlations between ratings of the Rorschach examiners and the criterion ratings. Here again the total correlation is positive and significantly greater than zero. The distribution of r's among the six judges is such as to yield a chi square of 6.77, $p = .30$ and we may again regard the distribution as relatively homogeneous.

Reference to Table 2 suggests that again suspected contamination apparently does not affect the results to a significant degree.

Rorschach factors reputedly related to "inner emotional tension" were studied. The first of these were the responses involving (1) human movement perceptions (M), (2) animal movement (FM), and (3) perception of inanimate objects in motion (m). The following correlation matrix was found in this investigation:

1 Criterion ratings
2 *M* responses
3 *FM* responses
4 *m* responses

$$r_{12} = -.27 \qquad r_{23} = .49 \qquad r_{34} = .22$$
$$r_{13} = -.04 \qquad r_{24} = .31$$
$$r_{14} = .11$$
$$M_1 = 5.52 \quad M_2 = 6.30 \quad M_3 = 4.36 \quad M_4 = 0.85$$
$$\sigma_1 = 1.14 \quad \sigma_2 = 5.11 \quad \sigma_3 = 3.21 \quad \sigma_4 = 1.17$$

The correlation observed between the criterion ratings and human movement responses alone is approximately the same as that observed between the Rorschach judges and the criterion (.24). Computing the multiple correlation coefficient yields $R_{1.234} = .35$ which appears to be somewhat better than that yielded between judges' ratings and the criterion.

Other patterns studied included the biserial r between patterns of $M \overset{=}{<} FM$ on the one hand and $M > FM$ on the other. This yielded practically no correlation ($r_{bis} = .09$). These same two scores were combined to give an index ranging from 0 to 2 as follows: $1-(M-FM)/(M+FM)$. The resulting indices correlated with the criterion ratings almost not at all ($r = -.06$). The ten patterns used with Variable 23b were used again in this instance, and yielded a Pearsonian r of .23; this is about as good as the correlation between the ratings of the judges and the criterion.

D. VARIABLE 26: SEXUAL ADJUSTMENT

The general question involved in this variable was: *"To what degree do his sexual needs and activities affect his over-all adjustment?"* Here, the lower the rating, the more it indicated a person whose "sexual needs and activities seriously interfere with his over-all adjustment." The higher the rating, the more likely it indicated an individual whose "sexual needs and activities definitely enhance his over-all adjustment."

Column four of Table 1 presents the findings with respect to the ability of the Rorschach examiners to predict the criterion ratings on sexual adjustment.

The results here are similar to those obtained on the previous three variables. The total correlation coefficient is positive and significantly greater than zero at less than the 1 per cent level. The distribution of r's yields a chi square $= 1.64$, $p = .90$. In line four of Table 2 we observe a significant difference between the r's at the 5 per cent level, but not in the expected direction for contamination to have been the cause, since the uncontaminated r is significantly greater than the contaminated one. There is no clear explanation for this type of difference in relation to the contamination hypothesis.

Rorschach scores selected for study in relation to sexual adjustment included (1) M responses, (2) FM responses, (3) the $M : FM$ ratio and (4) the factor designated as "sexual shock." Briefly, the first three scores and patterns yielded essentially no evidence of correlation being of the order of magnitude $-.09$, $-.02$, and $.05$ respectively. Biserial correlation, on the other hand, in which the dichotomous variables were absence of "sexual shock" and presence of "sexual shock" produced a point biserial r with the criterion ratings of .37 which is significant at less than the 5 per cent level. In using this result, one predicts a mean criterion rating on sexual adjustment of 4.33 for the group having little or no "sexual shock"; the best prediction for the "sexual shock" group is a criterion rating of 3.43. These means are significantly different at less than the 1 per cent level of confidence. Each of these means is also significantly different from the mean of the total group (3.82) at about the 1 per cent level.

E. VARIABLE 27: MOTIVATION FOR PROFESSIONAL STATUS

Involved in this variable are ratings aimed at stating how strong the subject's drive is for the status rewards of a professional career. Low ratings tend to indicate low drive, high ratings suggest strong drive.

Column five of Table 1 presents the findings indicating the ability of the Rorschach judges to estimate the criterion ratings on this variable. The result here, as reflected in the correlation coefficient of .32, indicates positive concomitant variation to a degree significantly greater than zero at better than the 1 per cent level of confidence.

Chi square for the distribution of judges' r's is 12.36, $p < .05$. This is the only array of r's which is apparently not homogeneous. In this instance the correlations of Judges B and *E are* homogeneous and so are those of Judges A, C, D, and F. The two groups of homogeneous r's were then combined by the z transformation and produced a difference significant at less than the 1 per cent level. This suggests the probability that Judges B and E were significantly better at estimating the criterion ratings than were the other four judges.

Line five of Table 2 gives the only evidence in favor of the hypothesis that there might be significantly greater correlation between the Rorschach ratings and the criterion contaminated by the Rorschach examiner's presence on the criterion team than between the Rorschach ratings and an uncontaminated criterion.

The series of Rorschach scores and patterns studied relative to this variable were: (1) number of responses utilizing the whole blot (*W*), (2) per cent of responses using the whole blot (%*W*), (3) total number of responses given (*R*), and (4) the normative score given by the formula (*W*-2*M*) yields values designed to show the relative excess of *W* over normal expectations of *M*.

All of these scores yielded negligible coefficients. That between *W* and the criterion was −.04; between %*W* and criterion −.07; between *R* and criterion −.12; and (*W*-2*M*) correlated with the criterion −.17. None of these are positive nor are they significantly greater than zero at the 5 per cent level of confidence.

F. VARIABLE 28: MOTIVATION FOR SCIENTIFIC UNDERSTANDING OF PEOPLE

Here the raters are asked to judge how strong the subject's drives are toward acquiring the facts, theories, and skills necessary for the scientific understanding of individual human beings. A low rating signifies low drive, a high rating a strong one.

Column six of Table 1 shows the findings obtained with respect to this variable. The total correlation is .38 and is both positive and significantly greater than zero at less than the 1 per cent level.

The distribution of *r*'s among the six judges yields a chi square of 0.995, $p > .95$. This may be interpreted as indicating a homogeneous distribution of *r*'s.

The study of the possible effect of a contaminated criterion yielded the results presented in Table 2. From these results it may be concluded that contamination probably had little or no effect in the direction hypothesized.

Rorschach factors studied in relation to "understanding people" were: (1) number of responses in which the content was human beings (*H*), (2) per cent of responses involving human beings (%*H*), and (3) number of responses in which percepts were of the type commonly seen by other persons (*P*). None of these produced correlations significantly greater than zero.

G. VARIABLE 29: INSIGHT INTO OTHERS

The task here was to rate the subjects on the amount of insight they possess into the attitudes, emotions, and motivations of others. A low rating suggests an individual who "interprets behavior at its face value; is insensitive to any but gross differences in behavior; does not develop any integrated understanding of behavior or of people." A high rating indicates the person who "has good awareness of underlying dynamics of behavior; is sensitive to subtle nuances of behavioral responses; is able to develop integrated understanding of the behavior of people."

The correlational findings relative to the ability of the "Rorschachers" to estimate the criterion ratings on Variable 29 are given in column seven of Table 1. Here again the total correlation of .23 is both positive and significantly greater than zero. The distribution of *r*'s among the six judges yields a chi square of 1.87 for which *p* is .90. This is interpreted as indicating a homogeneous array of *r*'s.

The contamination-study results on this variable are shown in Table 2. The null hypothesis that there is probably no difference between these *r*'s is retained, and the idea that contamination was a factor operating to increase the degree of correlation is rejected.

Two Rorschach scores were considered sufficiently relevant to this variable to warrant study: (1) Sum of responses in which form elements were primary and combined with textual and diffuse shading determinants ($Fc + FK$) and (2) Sum of responses in which form and color were combined determinants, form predominating (FC). The former values when correlated with the criterion ratings produced a coefficient of .16, the latter a coefficient of .17. While both of these are positive, they only approach the value (.18) which is significantly greater than zero at the 5 per cent level.

H. VARIABLE 30: INSIGHT INTO HIMSELF

On this variable the judges were asked to rate each subject as to "how much insight does he have into the underlying dynamics of his own attitudes, emotions, and motivations?" The higher the rating, the more the rater felt the subject had insight into himself.

The correlations obtained in connection with this variable are presented in column eight, Table 1. The correlation coefficient of .27 is positive and significantly greater than zero at less than the 1 per cent level. The distribution of coefficients among the six judges produced a chi-square value of 3.903; p is .60 and the distribution may be regarded as homogeneous.

The "contamination" study produced a significant difference but, as in the case of Variable 26, the direction of the difference is the reverse of that expected on the basis of contamination effects.

The same two Rorschach patterns which were used in the preceding variable were studied in relation to this variable: (1) ($Fc + FK$), and (2) FC responses. The resulting correlations were positive but not significantly greater than zero. The obtained r's were .08 for the pattern ($Fc + FK$) and .17 between FC responses and the criterion, the latter approaching significance at the 5 per cent level of confidence.

I. VARIABLE 31: QUALITY OF INTELLECTUAL ACCOMPLISHMENTS

The key question involved in this variable was: *"What is the characteristic quality of his intellectual output?"* A high rating suggests one who "characteristically produces intellectual work of high quality." A low rating, one whose "intellectual work is characteristically of low quality."

The correlation observed between the judges and the criterion on this variable was .33. The breakdown of correlations by judges is shown in column nine, Table 1. The total correlation is positive and significantly greater than zero. The distribution of r's among the judges produced a chi square of only 3.867 and p of .60 which indicates the probability that the distribution is homogeneous. Again there is no evidence that contamination had the effect expected.

The Rorschach scores studied in relation to "quality of intellectual accomplishments" included: (1) per cent of responses determined by "good" form (%F+), (2) per cent of responses to animal figures (%A), (3) number of original responses (O) and (4) number of good quality original responses ($O+$). In all of these, positive correlations are expected with the exception of %A which would be expected to be negative. The correlation coefficients obtained are given in Table 3. The correlations involving %A and $O+$ approach significance at approximately the 5 per cent level of confidence.

J. VARIABLE C-42: OVER-ALL SUITABILITY FOR CLINICAL PSYCHOLOGY

In the assessment rating scale, Scale C is concerned with characteristics more directly related to clinical psychology per se. It was felt that it would be of interest to include Item 42 for two reasons: (1) because it is the rating relative to over-all suitability, and (2) because it was used in the Preliminary Report as a criterion rating. The rating is future-oriented in that it required a rating "of the subject as he may be five years from now." The reference population is VA Clinical Psychologists who have had four years of academic and on-the-job training (U. S. Civil Service grade-classification of "P-4" or above[1]). The general question was worded thus: *"In view of his assets and liabilities, how well will he be able to carry out the several duties, diagnosis, therapy and research specified for the position of clinical psychologists (P-4 and above) in the VA?"* The higher the rating, the greater his suitability; the lower the rating, the less his suitability with the added stipulation that a "rating of 3 or below means that it is doubtful if the candidate will ever qualify for a P-4 rating."

The results of the correlation analyses for Variable C-42 are presented in column ten of Table 1. The total correlation of .35 is positive and is significantly greater than zero at better than the 1 per cent level of confidence. The distribution of *r*'s by judges is homogeneous; chi square being 5.384, $p = .40$. Here again contamination does not appear to be a factor operating to increase correlation (Table 2).

There are, of course, no Rorschach scores or patterns of scores designated as related to this variable but since it is an over-all rating, plausibly related to social adjustment, the ten "signs" employed in relation to social adjustment on Variable 23b were used again here. The obtained correlation coefficient is .52.

K. THE JUDGES

As noted above, there was some question of variability among the judges with respect to their ability to estimate the criterion ratings. This question was answered in part in the studies of the distribution of *r*'s among the six judges variable-by-variable. In brief, all of the distributions indicated the probability of homogeneity with the exception of that on Variable 27 where Judges B and E appeared to be able to predict to a significantly greater degree than the other four judges.

It is the writer's opinion that significant variability does, in fact, exist among the judges even though the above studies seem to contraindicate it. Partial justification for such an opinion may be seen in the following. The individual judges' predictions may be viewed from another standpoint, that of the number of variables each was able to predict to an extent significantly greater than zero at better than the 5 per cent level of confidence (reading across Table 1). This analysis results in the following information:

1. Judges A and B were each able to predict *six* of the ten variables at approximately the 5 per cent level or less.
2. Judge E was able to predict *five* variables at approximately this level.
3. Judge C was able to predict only three variables.
4. Judge D was able to predict only one variable.
5. Judge F was able to predict none.

---

[1] The grade-classification, "P-4," is now designated as "GS-11" in official VA and U. S. Civil Service regulations.

On a priori grounds, there being no statistical technique known to the writer for analyzing such data, it would appear that Judges A, B, and E might be regarded as better judges than C, D, and F, who among them made only 4 predictions at approximately the 5 per cent level or better, compared with the 17 significant predictions made by the former group.

It is obvious that using judges who lack the ability to make accurate predictions would operate as a systematic factor in lowering the over-all correlation between judges' ratings and the criterion. Stated differently, given six judges, all of the caliber of Judges A, B, and E, we should expect significantly higher over-all correlations with the criteria than those obtained and reported above. Other logical arguments in favor of the general hypothesis that the judges vary greatly in ability are evident in even a cursory examination of the experiential backgrounds of the judges themselves. Professional courtesy forbids stating here all of these evidences since to do so would leave open the possibility of identifying the judges and might reflect on them personally. These considerations suggest an ancillary investigation of the importance of the experiential backgrounds of Rorschach judges.

L. The Rating Process

Certain factors in the rating process itself were studied in an effort to try to understand the relative lowness of the correlation coefficients between judges' ratings and criterion ratings.

Among these were included such items as systematic errors of measurements (or better, errors in rating) with regard to the requirement that the raters fit their ratings into a predetermined distribution (see Appendix A); limited range or spread of ratings among the eight provided by the scales; and the setting in which ratings were made. Some evidence was uncovered to support the idea that the raters did not produce a "good fit" in distributing their ratings as prescribed and also that there was a limited range of ratings used. *While these errors may be trivial in their consequences,* they are reported as sources of error for what they are worth.

The fact that the Rorschach examiners found themselves functioning in a situation somewhat removed from their usual clinical settings, under a certain amount of pressure of time, and making unique interpretations in the form of ratings on a rating scale with which they had had limited experience is a factor worthy of consideration. At least this suggests that the general situation was not an optimum one for careful, accurate work with a complex instrument.

While nothing can be done about the first two factors at the moment, the matter of experience with the rating scale might suggest the hypothesis that with experience in the use of the scale, correlation might increase. This possibility was examined by dividing the six classes in half and comparing the correlations based on Rorschach estimates made during the first three weeks with those based on Rorschach estimates made during the last three. The results of this analysis are presented in Table 4. Here it will be observed that in general the correlations either remained about the same or increased. Thus, in Variables 24, 25, 26, and 27 there is no marked change; in the remaining variables there appears to be a more pronounced change *in the direction expected*. The analyses of the significance

TABLE 4

SIGNIFICANCE OF DIFFERENCE BETWEEN RORSCHACH AND CRITERION CORRELATION COEFFICIENTS FOR FIRST THREE WEEKS (I–III) AND LAST THREE WEEKS (IV–VI) OF ASSESSMENT

Variable	Obtained *r*		Diff.$_z$	Critical Ratio	*P*
	I–III	IV–VI			
23b	.25	.49	.2806	1.56*	.06
24	.43	.41	.0243	0.13	.45
25	.26	.23	.0320	0.18	.43
26	.29	.24	.0538	0.30	.38
27	.31	.35	.0449	0.25	.40
28	.38	.51	.1627	0.90	.18
29	.18	.30	.1275	0.71	.24
30	.16	.43	.2985	1.66	.05
31	.25	.44	.2168	1.21	.11
42	.24	.48	.2782	1.55	.06

* $\sigma_{D_z} = 0.18$ based on N (I–III) = 61 and N (IV–VI) = 59.

of the differences between these *r*'s reveals that the differences are probably "true" ones in the instances of Variables 23b, 30, and 42 at or about the 5 per cent level of confidence. Thus, it would appear that there is some reason to believe that experience with the rating scale led to better predictions.

Certainly in connection with all these considerations, mention must be made of the fact that correlations of this type are undoubtedly lowered by two other important factors: (1) the fact that the Rorschach is as yet only a crude instrument and even were it otherwise, one would not expect anywhere near a perfect analysis or evaluation through any single psychometric technique; and (2) the fact that the rater himself, making subjective estimates such as he is required to do, is no doubt a potent source of uncontrolled error of estimate.

CHAPTER IV

SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

THE problem of this study is the validity with which personality attributes may be estimated from data obtained by means of the Rorschach method of psychodiagnosis. Two aspects of the general problem have been considered: (1) the validity of estimates made by Rorschach examiners based upon Rorschach protocols they had obtained; and (2) the validity of various Rorschach scores and patterns of scores taken directly from the Rorschach records and regarded as related to the variables in question.

The procedure was to collect criterion ratings made by teams of observers who had had the opportunity to observe from many diverse viewpoints 120 male, VA trainee subjects throughout a week of intensive assessment activities provided in the University of Michigan Assessment of Clinical Psychologists Project during the summer of 1947. The subjects were assessed in groups of 24 over a period of six weeks. Assessment ratings were made on a specially devised rating scale at the end of each week of assessment. In this study only the more covert, genotypic variables constituting Scale B and the "over-all suitability" rating on Scale C were employed. Next, clinical ratings on the same variables made by Rorschach examiners and based on the Rorschach protocols and psychograms were obtained and correlated with the criterion ratings. Since the latter ratings were made by a total of six Rorschach examiners, the correlations between the ratings of each individual examiner and the criterion were also run in an effort to estimate the factors of ability and variability among the examiners. In each case, the total correlation based on all 120 cases was tested to determine the possibility of curvilinearity, with the result that all of the correlations appeared to be linear, *eta* never exceeding the product moment coefficient at less than the 5 per cent level of confidence.

For each of the 10 variables, Rorschach scores and patterns deemed to be of value in estimating the variable in question were isolated. These were then taken directly from the Rorschach records and correlated with the criterion ratings by correlational methods most in keeping with the type of data being studied; these included product-moment, biserial, and multiple-correlation techniques.

The results of these investigations indicated that in all 10 of the variables studied, the Rorschach examiners were able to estimate the criterion ratings positively and to a degree significantly greater than zero at approximately the 1 per cent level of confidence. The range of the correlation coefficients was from .23 to .43.

Since Rorschach examiners sat on half of the criterion teams, the possibility of their having contaminated the criterion ratings and thus having increased the correlations between their Rorschach ratings and the criterion ratings they helped to establish, was studied by comparing the "contaminated" correlations with the "uncontaminated." It was concluded that contamination as defined above was not particularly effective in spuriously increasing the correlation coefficients.

The several Rorschach scores and patterns studied revealed much variability in their capacity to predict the criterion ratings, ranging from low negative correlations to a positive .52. Some of this variability is undoubtedly due to the fact that in several instances the Rorschach patterns used were only approximations of the variables in question. Thus, Variable 27: Motivation for Professional Status was regarded (in terms of the Rorschach) as essentially a question of "drive and ambition"; Variable 28: Motivation for Scientific Understanding of People was interpreted as "interest in and empathy for others"; and the two "insight" variables, 29 and 30, were correlated with Rorschach patterns suggestive of "insight into anxieties," rather than insight into others and into himself as defined on the rating scale. Correlations significantly greater than zero were obtained on the following variables:

1. Variable 23b: Social Adjustment
2. Variable 24: Appropriateness of Emotional Expression
3. Variable 25: Characteristic Intensity of Inner Emotional Tension
4. Variable 26: Sexual Adjustment
5. Variable 31: Quality of Intellectual Accomplishments
6. Variable C-42: Over-all Suitability for Clinical Psychology (defined as "social adjustment" in selecting Rorschach patterns).

In comparing the obtained correlations in the above instances with the correlations between the clinical ratings made by the Rorschachers and the criterion ratings, it was found that in no case was there a difference significant at the 5 per cent level or less.

In all the other variables, the clinically oriented estimates were superior, and it is interesting to note that these are precisely the variables noted above for which only approximate Rorschach patterns could be devised.

Since it is commonly felt that correlations should be near .60 in order to speak of valid predictions, it would appear that the obtained results fall far short of indicating validity for either the ratings made by the Rorschach examiners or the estimates based on Rorschach scores. On the other hand, when investigating psychological problems, even very small correlations, if statistically significant, are frequently indicative of psychologically lawful relationships. The fact that the correlation coefficients are small may only mean that the measurement situation is contaminated, uncontrolled, not clearly defined, or that measurement itself is crude owing to the measuring instrument employed. In the present study, it has been demonstrated that many of these latter factors were operating to varying degrees.

While there was in general no statistically significant degree of variability among the predictions of the six Rorschach judges on the several variables, evidence in the form of the number of variables estimated at less than the 5 per cent level of confidence indicates the probability that three of the Rorschach judges were relatively "good," producing a total of 17 significant predictions, in contrast to the remaining three who produced a total of only 4 significant estimates. It is probable that the three inferior Rorschach raters tended to lower the total correlations. Partial evidence would support this split of the judges on the basis of their professional and experiential backgrounds.

In view of the foregoing, and in view of the relative inadequacy of any single technique such as the Rorschach method to supply all the answers, it is not surprising that the observed correlations were as low as they were, and gratifying that they were so consistently positive and statistically significant. It is the writer's opinion that had a large part of the sources-of-error variance been controlled, correlations much nearer to a level suggestive of valid prediction would have resulted.

On the basis of these results we may conclude that:

1. Clinicians, using Rorschach data, are able to make estimates of covert personality characteristics with a validity coefficient of the order of .35, as represented by the present sample of ten attributes.

2. Relatively clear personality attributes and complexes, such as "social adjustment" and "appropriateness of emotional expression," can be equally well estimated from certain Rorschach scores or patterns of scores; those which are more specific, such as "motivation for professional status" or "motivation for scientific understanding of people," were not as successfully estimated by the score approach as they were by the clinical approach.

3. Some Rorschach examiners may be adjudged "good" raters while others must be regarded as "poor" raters; the difference seems to lie in the difference of their backgrounds of professional training, experience, and interests.

A number of problems for further investigation suggest themselves, a few of which may be mentioned here. In the first place, a new set of ratings, completely "blind," might be obtained from a new panel of Rorschach "experts," in order to evaluate the possible effects of personal contact with the subjects. Second, the problem of reliability of Rorschach ratings should be studied by having a number of Rorschach examiners rate the same records and by studying interjudge agreements. Third, more extensive study should be made of the importance of extent of familiarity with the Rorschach method, breadth of general clinical experience, working knowledge of personality dynamics, and so forth as factors correlated with the ability of clinicians to estimate personality attributes from data obtained by the Rorschach and other projective methods. Fourth, and related to the preceding suggestion, work should be done to determine the more specific effects of the personality of the clinician upon his Rorschach-based ratings of personality attributes.

APPENDIX A

RESEARCH PROJECT ON THE SELECTION OF CLINICAL PSYCHOLOGISTS

REFERENCE POPULATIONS FOR RATINGS

For Scales A and B the candidate is rated as he is today. The reference group is first-year clinical psychology graduate students in universities accredited by the APA to offer training in clinical psychology.

For the Criterion Skills (Scale C), the candidate is rated as he will be five years from now. The reference population is VA clinical psychologists who have had four years of academic and on-the-job training (P-4 and above). In rating a candidate on the Criterion Skills, the assumption is made that he will have completed the training and will have been employed by the VA. A rating of 3 or below means that it is doubtful that the candidate will ever qualify for a P-4 rating with respect to the trait being rated. It is possible that some candidates will later qualify for a P-4 position even though they rate 3 or below on a few Criterion Skills.

RATING SCALE

All ratings will be made in terms of an eight-point scale illustrated graphically in Figure 1.

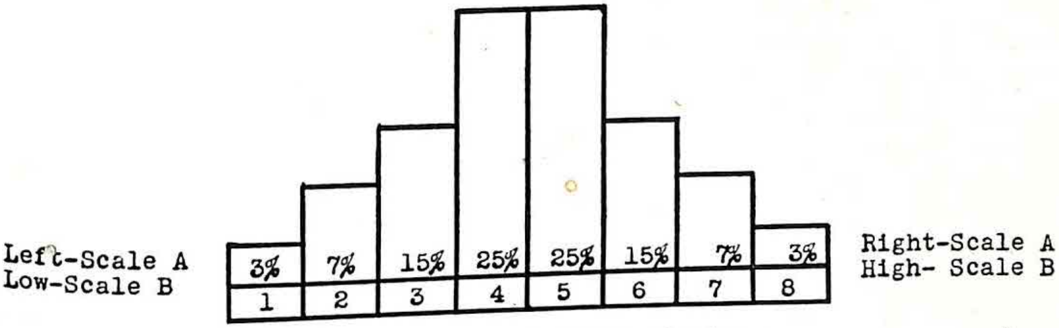
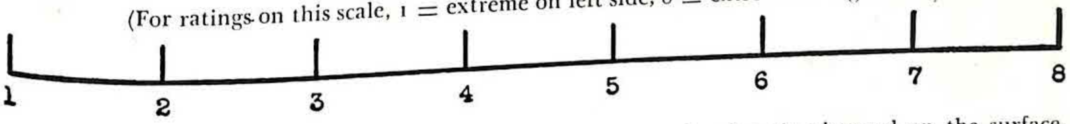


FIG. 1. Theoretical distribution of ratings

RATING SCALE DEFINITIONS

Scale A (# 1-22)

(For ratings on this scale, 1 = extreme on left side, 8 = extreme on right side.)



Note: The attributes in Scale A refer to behavior which can be directly observed on the surface. In using this scale, disregard any inferences about underlying dynamics or causes.

1. *Readiness to cooperate* Finds ways of cooperating despite difficulties.	–v– *Obstructiveness* Frequently is "difficult."
2. *Consistent* Behaves in the same general way from day to day.	–v– *Inconsistent* Shows changing and unpredictable moods and behavior.
3. *Assertive* Attempts to dominate or influence his associates.	–v– *Submissive* Tends to let people have their way.
4. *Depressed* Does not smile or laugh easily or frequently.	–v– *Cheerful* Generally bubbling over with good cheer. Optimistic. Enthusiastic. Prone to cheerful, witty remarks.

5. *Irresponsible* Not inclined to take responsibilities seriously. Thoughtless. Unaware of responsibilities of his age. (Do not confuse with #17: Conscientious vs. not conscientious.)	–v– *Serious* Accepts appropriate responsibilities toward others. Shows seriousness of purpose.
6. *Gregarious (adient toward people)* Is attracted by and moves toward people in social situations.	–v– *Non-gregarious (abient from people)* In social groupings, isolates himself.
7. *Easily Upset* Easily embarrassed or put off balance. Gets confused in emergency. Blushes, shows excitability, becomes incoherent. Momentary "nervousness," *not* general emotionality.	–v– *Unshakable* Self-possessed. Does not lose composure under emotional provocation.
8. *Narrow Interests* Uninformed in many areas. Narrow, simple, interests. Provincial outlooks.	–v– *Broad Interests* Talks and acts *in an informed way* in a wide variety of areas.
9. *Suspicious* Believes rather too quickly that he is being unfairly treated. Imagines on insufficient grounds that people strongly dislike him. Interprets things as having reference to himself when none is intended. Feels persecuted.	*Trustful* Accepts statements of others without suspicion (not necessarily "gullible.")
10. *Generous* Generous and considerate in actions involving others at considerable self-sacrifice.	–v– *Self-centered, selfish* Irritable and resentful or withdrawing from situations requiring consideration of others. Talks and acts exclusively in terms of own interests.
11. *Silent* Says very little.	–v– *Talkative* Talks a lot, to everybody. Takes the initiative in conversations. When addressed, responds quickly.
12. *Cautious* Avoids the strange and new. Looks at all aspects of a situation overcautiously. Does the safe thing.	–v– *Adventurous* Seeks and readily enters into new experiences and situations.
13. *Socially poised and adept* Polite, poised, and tactful in social situations. Deals with people gracefully and skillfully.	–v– *Socially clumsy, awkward in social situations* Tactless in social situations. Crude in speech and manners. Omits proper formalities. Does not meet people gracefully. Gauche. Note: Applies to relationships with one or more people.
14. *Rigid* Sticks to his own ideas and does not adapt to ways of doing things differently from his own. Does not change and broaden with experience.	–v– *Adaptable, flexible* Appropriately modifies his behavior to situations. Accepts compromises where needed. Is not upset, surprised, baffled, or irritable if things are different from what he expected.
15. *Dependent* Asks for reassurance and support. Attaches himself to individuals and groups instead of relying on himself.	–v– *Self-sufficient* Does not lean on others in situations calling for independent action.
16. *Placid* Outwardly calm and relaxed.	–v– *Worrying, anxious* Appears to worry constantly, harried; overtly anxious. Shows agitation.

17. *Conscientious* –v– *Not conscientious*

Behavior characterized by truthfulness, honesty, unselfishness.

Not too careful about right and wrong where own wishes are concerned. Not particularly just, honest, or unselfish. Inclined to somewhat shady transactions.

18. *Imaginative* –v– *Unimaginative*

Has a rich and vivid imagination. Thinks of unusual angles and aspects of a question. Sensitive to a multitude of emotional and other possibilities not realized by the average person.

Lacks imagination. Approaches problems in a literal matter-of-fact fashion. Unresponsive to the subtleties in a situation.

19. *Marked Overt Interest in Opposite Sex* –v– *Slight Overt Interest in Opposite Sex*

Dates a good deal and/or talks a lot about opposite sex. Extremely aware of women as women. (Disregard *inferred* needs or drives.)

Talks very little about women. Does not use opportunities for contacts with women.

20. *Frank* –v– *Secretive*

Comes out readily with his real feelings on various questions. Expresses his feelings, sad or gay, easily and constantly.

Keeps his thoughts and feelings to himself.

21. *Dependent Minded* –v–*Independent Minded*

Intellectually dependent on others. Generally accepts the opinion of a group or of authority without much thought.

Thinks things out for himself. Examines every question persistently and individualistically. Makes up his own mind.

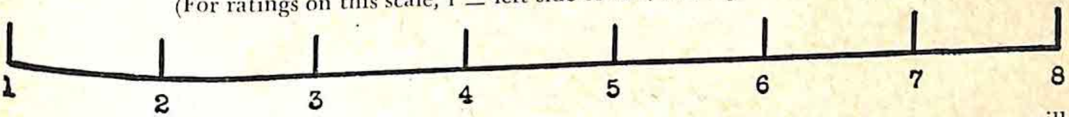
22. *Limited overt, emotional expression* –v–*Marked overt emotional expression*

Is apathetic, sluggish, or constricted.

Shows hyperkinetic, agitated behavioral responses; is overly excitable and over-demonstrative.

Scale B (# 23-31)

(For ratings on this scale, 1 = left side or low, 8 = right side or high.)



Since many of the following attributes (#23-31) are broad factors, it is unlikely that any person will fit all of the phrases grouped together at one pole of a given variable. Note also that for some items, neither extreme necessarily represents a desirable attribute.

23a. *Ability to develop and maintain warm interpersonal relationships.*

23b. *Social Adjustment:* How well does he adjust to varied interpersonal situations? (includes sexual adjustment only as it affects social adjustment in general.)

Acts without consideration for feelings of others; often rejected by others, often appears aloof, hostile, or irritable.

Actively considers feelings of others; readily gains acceptance in interpersonal relationships; maintains a friendly and likeable manner.

24. *Appropriateness of Emotional Expression:* How appropriate are his emotional responses to the situation?

Fails to adapt his emotional responses to the needs of the situation; shows disorganized or overly constricted emotional responses.

Shows emotional responses of a quality and intensity befitting the situation; reacts spontaneously but appropriately; shows well-integrated and flexible patterns of emotional behavior.

25. *Characteristic Intensity of Inner Emotional Tension:* How intense is his inner emotional life as *inferred* from all available clues?

Inner emotional life characterized by a minimum of persistent internal tensions.

Has strongly repressed emotional drives resulting in inner turmoil; great inner conflict and strong pent-up emotions.

26. *Sexual Adjustment:* To what degree do his sexual needs and activities affect his over-all adjustment?

His sexual needs and activities seriously interfere with his over-all adjustment.	His sexual needs and activities definitely enhance his over-all adjustment.

27. *Motivation for Professional Status:* How strong is his drive for the status rewards of a professional career?

28. *Motivation for Scientific Understanding of People:* How strong are his drives toward acquiring the facts, theories, and skills necessary for the scientific understanding of individual human beings?

29. *Insight into Others:* How much insight does he have into the attitudes, emotions, and motivations of others?

Interprets behavior at its face value; insensitive to any but gross differences in behavior; does not develop any integrated understanding of behavior or of people.	Has good awareness of underlying dynamics of behavior; is sensitive to subtle nuances of behavioral responses; is able to develop integrated understanding of the behavior of people.

30. *Insight into Himself:* How much insight does he have into the underlying dynamics of his own attitudes, emotions, and motivations?

31. *Quality of Intellectual Accomplishments:* What is the characteristic quality of his intellectual output?

Intellectual work is characteristically of low quality.	Characteristically produces intellectual work of high quality.

Scale C—Criterion Skills (#32-42)

(For ratings on this scale, 1 = low, 8 = high)

Ratings on #33-42 refer to the subject's performance five years hence.

What will be his level of competence or skill in the varied aspects of:

32. *Academic Performance:* (during next three or four years) How well will he:
Effectively master course content, successfully complete courses in general psychology, clinical psychology, statistics, and related fields; satisfy language requirements for the doctorate; pass general examinations.

33. *Clinical Diagnosis:* How well will he:
Recognize dynamics underlying particular responses in both objective and projective tests, observe significant interrelationships among responses, relate findings to case history and other clinical data.

Elicit from the patient information required for mental status examinations and case histories; ascertain and evaluate attitudes and incidents of psychological significance in the patient.

Synthesize clinical findings to arrive at an integrated picture of personality development, structure, and function.

34. *Individual Psychotherapy:* How effectively will he:
Conduct various types of individual psychotherapy.

35. *Group Psychotherapy:* How effectively will he:
Conduct various types of group psychotherapy.

36. *Research:* How well will he:
Recognize and define important research problems in clinical psychology; critically evaluate and apply the research findings of others; think with originality and scientific rigor; employ appropriate experimental design and statistical methods; grasp practical implications of findings, present results and conclusions in clear, comprehensive, and well-organized form.

37. *Administration:* How well will he:
Plan and develop psychological programs; make proper administrative decisions; delegate responsibility appropriately; elicit cooperation from subordinates and superiors; maintain high morale among his staff; carry out or direct an appropriate public relations program.

38. *Supervising Clinical Psychologists:* How well will he:
Carry out the professional supervision of subordinates assigned to him for duty and on-the-job instruction; assign their duties; evaluate their performance; instruct them in clinical techniques; perform other aspects of in-service training.

39. *Teaching Psychology* (in a College or University): How well will he:
Teach college courses in general psychology: motivate students, present concepts and procedures, stimulate critical thinking about and integration of course materials, evaluate the products of learning.

40. *Professional Interpersonal Relations:* How well will he:
Work cooperatively with superiors, subordinates, members of the mental hygiene team, and other professional personnel concerned with the patient's welfare; participate in the give and take of staff conferences; contribute to group decisions.

41. *Integrity of Personal and Professional Behavior:* How well will he:
Recognize and fulfill professional responsibilities; live up to personal commitments; show loyalty to professional obligations in the event of outside pressure or promise of personal gain; maintain discretion concerning professional matters; appropriately conform with commonly accepted standard of moral and social behavior; refrain from coloring facts, evasion, lying, etc.

42. *Over-all Suitability for Clinical Psychology:* In view of his assets and liabilities, how well will he be able to:
Carry out the several duties, diagnosis, therapy, and research specified for the position of clinical psychologist (P-4 and above) in the VA.

REFERENCES

1. BECK, S. J. *Rorschach's test, Vol. I: Basic processes.* New York: Grune and Stratton, 1944.
2. CATTELL, R. B. *Description and measurement of personality.* Yonkers-on-Hudson: World Book Company, 1946.
3. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique.* Yonkers-on-Hudson: World Book Company, 1942.
4. MIALE, F. R., & HARROWER-ERICKSON, M. R. Personality structure in the psychoneuroses. *Rors. Res. Exch.*, 1940, 4.
5. RIDER, P. R. *Modern statistical methods.* New York: Wiley, 1939.
6. Progress Report and Preliminary Findings: University of Michigan project in the assessment of clinical psychologists. Mimeographed and privately circulated, 1948.

VOLUME 66
NUMBER 7

WHOLE NO. 339
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Development of an Eye Camera for Use with Motion Pictures

By

PAUL R. WENDT
San Francisco State College

with Foreword by

WALTER W. COOK
University of Minnesota

Accepted for publication, April 6, 1951.

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

FOREWORD

Although there has been considerable research on the over-all effectiveness of educational motion pictures, very little is known about their psychological impact on individuals. We know very little about how we communicate meanings by motion pictures. The design of film scenes and sequences is left largely to the intuition of producers. Investigations of perception have covered some of the basic considerations in how meaning is conveyed by visual teaching materials but many experiments are necessary before we get sufficient clues as to how best to construct these materials. New techniques and new apparatus need to be developed for this research.

This monograph describes one approach to this problem. The eye camera used previously in research in eye movements in reading was adapted for use with motion pictures. The camera is synchronized with the educational motion picture being viewed so that fixations can be determined for each frame of the teaching film. The eye-camera film moves intermittently, making a framed motion-picture record. This can be superimposed directly on the educational motion picture viewed by the subject. Eye dots can be located directly on the motion-picture screen indicating exactly each eye fixation. Corrections are easily made for slight head movements. Thus we can quickly determine what a student looks at when viewing a motion picture.

This equipment and improvements on it may make it possible for us to make analyses of how we learn from motion pictures and how they could be better designed to teach more effectively.

Walter W. Cook

University of Minnesota
December 1, 1950

TABLE OF CONTENTS

Foreword iii

I. *Introduction* 1

II. *Criteria for a method of measuring eye movements of subjects viewing motion pictures* 1

III. *Methods of making eye records* 3

IV. *Adaptation of the Brandt camera* 4

V. *The making of eye-movement records in the laboratory* 11

VI. *Analysis of the eye-movement records* 13

VII. *Summary* 17

Bibliography 18

DEVELOPMENT OF AN EYE CAMERA FOR USE WITH MOTION PICTURES*

I. INTRODUCTION

Motion pictures have come into wide use in the field of teaching and are destined for much more use. But experimentation has, so far, merely indicated that they are effective, not why they are effective nor how this effectiveness is achieved. Furthermore, research has been concerned entirely with the over-all effectiveness of an entire film, not with the effectiveness of scenes or sequences within the film. Yet the teaching value of an entire film is based on the teaching value of its parts. Within any one motion-picture scene, it is still impossible to tell which elements are effective in teaching and which are not.

The determination of the effectiveness of parts of films will be of interest primarily to educators who wish to exert more control over the translation of their ideas into motion-picture scenes. No one can tell, furthermore, whether the money and effort put into any one scene by the producer is justified by the actual learning accomplished in a classroom where the film is used. If such information were available, large savings might be possible.

Methods need to be devised to determine which elements contribute to the teaching value of the film and which do not. One line of attack is to adapt the eye camera used in measuring eye movements in reading to measuring eye movements while pupils are looking at educational motion pictures. This paper describes this method.

II. CRITERIA FOR A METHOD OF MEASURING EYE MOVEMENTS OF SUBJECTS VIEWING MOTION PICTURES

The eye camera had to satisfy the following criteria:

1. It must provide a record of both vertical and horizontal eye movements on one film.

2. It must move the film intermittently as in a motion-picture projector, so that the resulting record can be projected directly for analysis.

3. It must operate as nearly as possible at the projection speed of educational motion pictures (24 frames per second).

4. It must provide maximum comfort for the subject so that the situation of viewing the teaching film will be as nearly natural as possible.

A. Types of Eye-Movement Measures

Obtaining eye-movement measures of subjects looking at motion pictures differs in several important ways from obtaining eye-movement measures during the process of reading.

Not all the measures of eye movements which are useful in reading seem to be applicable to the viewing of motion pictures. The measures of eye movements are: (*a*) pause duration, (*b*) span, (*c*) fixation frequency, and (*d*) regression frequency. Tinker (6) has stated that of these, the two most closely associated with reading ability are fixation frequency and regression frequency, and that pause duration is one of the worst measures of reading ability because it varies so much with the type of reading material.

---

* The apparatus herein described was developed as part of a study conducted under the generous guidance of Dr. Walter W. Cook. Dr. Miles A. Tinker contributed from his extensive experience in analyzing eye movements. The original eye camera that is here adapted for motion pictures was developed by Dr. Herman R. Brandt. Mr. Peyton Stallings, Miss Dolores Paul, and Mrs. Evon Unruh contributed invaluable assistance.

Regression frequency cannot apply to pictures, either still or motion, because pictures are not read from left to right or in any standard direction as in reading, even though Brandt (1, p. 57) has found a slight tendency for the eyes to make more fixations in the upper left quadrant in viewing symmetrical designs.

Fixation frequency would probably not be a very good measure of eye movements with motion pictures because of the constant speed factor in motion pictures. Sound films move at a speed of 24 individual frames per second, and the average motion-picture scene lasts only 6 seconds. In a good teaching film, individual scenes follow each other in very rapid succession, putting pressure on the viewer to assimilate as much detail as possible in the time allowed, which is always too little to give the viewer time to see everything. In this situation, fixation frequency would probably be negatively correlated to a high degree with pause duration. Its value might be close to a mere inverse of the value of pause duration and merely duplicate the investigation of pause duration. Thus, fixation frequency theoretically seemed to have little possibility as a good measure of eye movements in this experiment, although it might prove to have some value after the first few pioneering experiments have been made to explore the field.

The two measures that remain, span and pause duration, seemed to be the best to use in this experiment. Although pause duration, as mentioned above, has turned out to be one of the worst measures of reading ability, it seemed on a theoretical basis that it might be much more useful as a measure of eye movements in scanning motion pictures. The term "span" in reading research commonly refers to the amount of one line of type which is comprehended by the reader in one fixation by the use of peripheral as well as foveal vision. Such a concept in viewing pictures would not be useful in these early stages of research on pictures. A comparable measure, however, would be *the linear distance in any direction traveled by the eye between any two consecutive fixations.* It was decided to call this distance the *interfixation movement,* to distinguish it clearly from span.

Mention should be made here of the term "return sweep" which is used in reading research. It refers to the movement of the eye in moving from one end of a line of type which has been read to the first fixation on the next line of type. Because the term return sweep has such a specialized meaning connected with the lines of type, it was decided not to apply it to viewing pictures.

The method of recording eye movements, therefore, should be capable of giving fairly exact measures of *interfixation movement* and *pause duration.*

B. NEED OF LOCATING FIXATIONS IN TWO-DIMENSIONAL PLANE

The measuring method needed also to tell *where the eye is fixated* in a two-dimensional plane perpendicular to the line of vision. In reading research, it is usually sufficient to measure the lateral movement of the eyes from left to right. In viewing pictures, it is necessary to locate these fixations on a two-dimensional plane by measuring not only the lateral movements, but also the vertical movements, simultaneously.

C. EYE-CAMERA SPEED AS NEAR AS POSSIBLE TO TWENTY-FOUR FRAMES PER SECOND

If an eye camera were used for measuring the viewing of motion pictures, it would be especially useful if the film advanced through the eye camera at 24 frames per second, the same speed that the motion picture is projected on the screen. If at this speed each fixation were represented by one dot in two dimensions in a frame of the film, then the eye-camera film would have a one to one relationship to the teaching film being viewed. It would be possible, [illegible] this ideal situation, to project the eye-camera film on the screen, thus superimposing the actual fixation dot on the motion-picture scene which caused this fixation.

It seemed at the beginning that it might even be possible, ideally, to project the eye-camera record in some instances by a motion-picture projector, superimposed on the screen on the teaching film image which, of course, would also be run in synchronism in a motion-picture projector. The effect on the screen would be that of a moving fixation dot, moving about on the screen exactly in the same way in which the subject viewed the teaching film. If such a method could be developed, it would even be possible to synchronize several projectors and contrast or compare the fixations of several subjects selected as typical or atypical from a particular point of view. In practice it was found that this ideal of making eye records at 24 frames per second and projecting them in a motion-picture projector could not be realized owing to the limitations of cameras now available. A compromise speed had to be accepted, and the motion-picture projection of eye records deferred to a later experiment.

D. Comfort of the Viewer

Another very desirable criterion of a machine for measuring eye movements with motion pictures was the comfort of the viewer. The basic reason for this was that printed words are symbols, whereas motion pictures are close to reality. Reading is, therefore, expected to be done in a somewhat constrained physical position, whereas the viewing of motion pictures allows considerable freedom of movement.

It was also necessary to make the eye-movement records in a room with subdued light, since the subjects would be viewing motion pictures. To view motion pictures in a room even partially illuminated would ruin any effective realism which the motion picture might be trying to create.

Finally, it was desirable to find a method of making eye-movement records which would not involve shining too bright a light into the subject's eyes, because this would be especially trying when the subject was looking at motion pictures in a darkened room.

III. METHODS OF MAKING EYE RECORDS

The standard corneal-reflection type of eye camera produces a record on 35-mm. film consisting of a series of dots closely spaced. The film is not moved intermittently as in a motion-picture camera but continuously as in a smoked-paper tracing.

Buswell (2) measured eye movements in two dimensions during viewing of pictures by using two recording films at right angles to each other moving synchronously and continuously. Analysis was made by computation.

The camera developed at Purdue University is simply an 8-mm. motion-picture camera taking pictures of the subject's eyes reflected in a half-silvered mirror. Trained analysts can tell at which large area of the material the subject is looking. The method lacks the exactness of the corneal-reflection type of camera but is quick and useful for such tasks as analysis of effectiveness of advertising layouts.

The camera developed by Dr. Herman F. Brandt uses a 35-mm. motion-picture camera moving the film intermittently to photograph the corneal reflection. It pro-

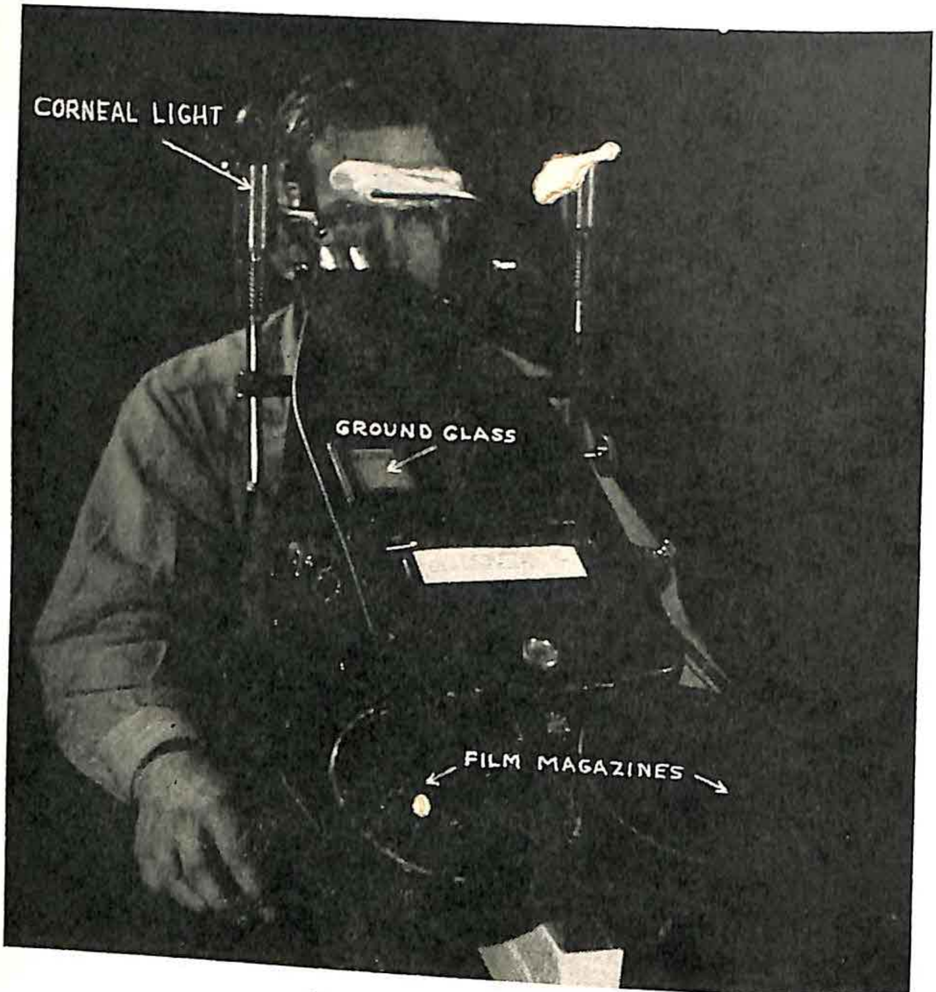


FIGURE 1. The eye camera.

duces a two-dimensional record on framed film. Head movements are also recorded on the film.

This camera came closest to meeting the criteria. It makes a framed motion-picture record which can be analyzed by projection in a filmstrip projector without computation. It has the accuracy of the corneal reflection type. It does not need bright photographic lights.

IV. ADAPTATION OF THE BRANDT CAMERA

A NUMBER of minor changes in the Brandt camera and some major changes in the equipment used with it were made to make it more useful in measuring eye movements of subjects looking at motion pictures.

A. INCREASE IN CAMERA SPEED

The normal speed of the Brandt camera is 2 frames per second. It was the original plan to devise some method of eye photography which would take eye records at 24 frames per second so that the resulting picture could be projected in a motion-picture camera, superimposed on the screen over the teaching film. It was clear that the Brandt camera mechanism could not be speeded [illegible] 12 times. The mechanism would not stand the strain.

It was possible, however, to speed the camera up to some extent. The motor drive of the camera was synchronous, and the motor itself contained a 3 to 1 reduction gear. By removing this gear and attaching the motor drive directly to the secondary gear box, the camera speed was increased from 2 frames per second to 6 frames per second. This was the maximum speed possible with this camera, and it was decided to settle for this, as such a speed would be ample to register most fixations with even the shortest possible duration.

B. Increase in Candle Power of Corneal Light

The corneal light sources were mounted in lamp houses on brackets on either side of the camera. A beam of light aimed at the eye was formed by a fine horizontal slit in the lamp housing. Only the right-hand light was actively used in the experiment since it illuminated both the cornea of the right eye and the surface of the false eye near the right temple. The left-hand corneal light was beamed at the left eye because it was more comfortable for both of the subject's eyes to be illuminated approximately evenly rather than to have one in darkness.

The camera came equipped with a corneal light source of a 1,183 headlight, producing a light intensity of 30 candle power. When the camera was speeded up, it was necessary to obtain more light to expose the photographic film, since the tripling of the camera speed reduced the exposure by one-third. Therefore, the right-hand corneal light was replaced with a 50 candle-power bulb of the same type. The left-hand corneal light source was left at 30 candle power since the imbalance in illumination on the two eyes was not noticeable to the subject.

C. Change of Type of False Eye

As mentioned previously, the camera makes provision for registering on the film the reflection from a false eye. This was a chromium bead attached to a pair of empty spectacle frames. Preliminary tests showed that approximately 10 minutes would be required to fit the spectacle frames to each of the subjects as they varied in physiognomy; therefore, the frames were replaced with a very light earphone head clip to which was attached a chromium cap screw. This cap had greater curvature than the chromium cap on the spectacle frames, and therefore produced a smaller false eye dot on the motion-picture film. It was attached to the earphone head clip by two inches of lead wire which was soft enough to be bent easily. Thus, it was easy to adjust the false eye into the optimum position for each subject. Subjects with broad skulls often required the bending of the false eye toward the right eye so that they would not be too far apart to register on the ground glass of the camera at the same time.

D. Adaptation of Viewing Angle to Motion Pictures

Dr. Brandt's camera was supplied with a mount for copy approximately 2½ feet in front of the subject's eyes and parallel to the plane of the motion-picture film. When the camera was used for this purpose, its natural position allowed the subject to look downward at an angle of approximately 30 degrees to look at the copy.

Even though it would have been possible to project the pictures on the advertising copyholder simply by using a white sheet of paper as a screen and focusing the projector for a very short throw, it was decided not to project the teaching film on a very small screen. When motion pictures are viewed in a reduced size at close range, even though the angle of view for the eyes is exactly the same as it is for a larger screen further away, the realistic effect of the motion picture is lost, probably because the necessity of focusing the eyes at short range makes the subject fully aware that he is not looking at even a close imitation of reality. When motion pictures are projected on a small screen, they give the impression of a peep show. On the other hand, if pictures are projected at a comfortable distance of 10 feet or more away from the subject, he is able to get somewhat of an impression that he is looking at a real scene and project himself into the picture. The Society of Motion Picture Engineers (5) has set up well-established standards for screen size in relation to the viewer in any room used for the screening of motion pictures. The standard states that no seat in the auditorium should be nearer to the screen than twice the width of the screen and no seat should be farther from the screen than five times the width of the screen.

Motion pictures are normally projected so that the subject's line of sight is approximately 10 degrees above the horizontal. To adapt the camera, therefore, for viewing motion pictures, the magazine rack was removed, the camera was elevated about a foot above the height in which it was normally used, and it was tilted down toward the subject. The front half of it was hinged to the table, in order to provide stability. Un[illegible] the back of the camera, a four-inch wooden wedge was placed. By inserting or withdrawing this wedge, the angle of the camera could be adapted to the individual subject. The result was that the reflected beam of light from the cornea entered the lens tube of the camera more horizontally and at about the same angle of reflection as it did formerly.

By extending the sight lines roughly from the subject's eyes past the magazine rack to the wall, the comparable size of the screen was determined. It was reduced somewhat to be safely within the standards of the Society of Motion Picture Engineers. It was not reduced more than absolutely necessary because the camera was designed so that the maximum movement of the subject's eyes in swinging from one side of the copy table to the other was within, and yet filled completely, the area of the eye-camera motion-picture frame. When the size of the screen had been determined, it was checked by actually placing a subject in front of the eye camera and turning on the corneal lights, positioning and focusing the camera, and then asking the subject to sweep his eyes to each of the four corners of the screen in succession while the experimenter checked the movement of the fixation dot on the reflex ground glass of the camera. The movements of the fixation dot were found to be just within the limits of the ground-glass aperture. The final size of the screen was 33″ x 44″ and the distance from the subject to the screen was 8 feet.

In order to keep the subject as much at ease as possible, plenty of foot room was provided by placing the camera on a small 30″ x 40″ table which provided clearance for the knees. As the subject was to be photographed in a slightly reclining position, the ability to extend his legs forward contributed considerably to his comfort. Since the camera was quite narrow, the subject's arms were placed on the sides of the table on either side of the camera in a relaxed position.

E. Use of an Adjustable Chair

Preliminary tests with the camera showed that, if the subject merely pressed his head against the curved support provided for the forehead and against the brace provided for the chin, the eye records showed excessive head movement. In spite of a desire to give the subject all the comfort possible in order to get as nearly a natural reaction to the motion picture as could be obtained, it was decided that some sort of headrest must be provided. At the same time, it was clear that either the subject's chair or the camera would have to be mounted on a lifting device in order to provide quickly the close vertical adjustments when each student was first seated. (The fine adjustments were made in the camera itself.) The ideal subject chair was found in a standard dentist chair operated by hydraulic lift. This type of chair has so many adjustments on it that it was possible to seat the subjects quickly and exactly in front of the lens tubes of the camera. The arms of the chair had been removed so that it could be slid under the table. Once the subject was seated, the back rest of the chair was brought up to provide a firm support for the subject's back and to make him sit upright so that it was impossible for him to change the level of his eyes by slumping in the chair. Then the hydraulic lift was used to bring the subject's eyes up to the level of the lens tubes so that the narrow beam from the corneal light was exactly centered on the cornea of the right eye. With the hydraulic lift, a very fine adjustment was possible so that the corneal light did not have to be adjusted more than a fraction of an inch.

F. Installation of Bite Stick

In the matter of the headrests, several tests were made. The headrest attached to the camera included both the chin rest and a place to insert a bite stick. Following the principle of using the least constrained position on the subject, some eye records were made with subjects merely resting their chin in the chin rest and pressing their forehead in the forehead rest. It was quickly found that the chin rest forced subjects with long chins to keep their heads higher in the headrest than normal, with the result that the curved band of the forehead rest sometimes came down over their eyebrows, partly obscuring their vision. Subjects with short, round faces found that the forehead rest (which was not adjustable) came too far up on the forehead to provide firm support. Even in small samples of subjects, it was inevitable that different face shapes would be encountered. Therefore, the chin rest was abandoned. The only alternative was to use the bite sticks. These were round, wooden sticks ⅛ inch in diameter and approximately 6 inches long, obtained from a candy factory. A supply of them was sterilized and kept on hand, a new one being used, of course, for each subject. They were inserted into the machine between two metal supports. Although it seemed unfortunate to introduce such an artificial situation as biting on a stick while viewing motion pictures,

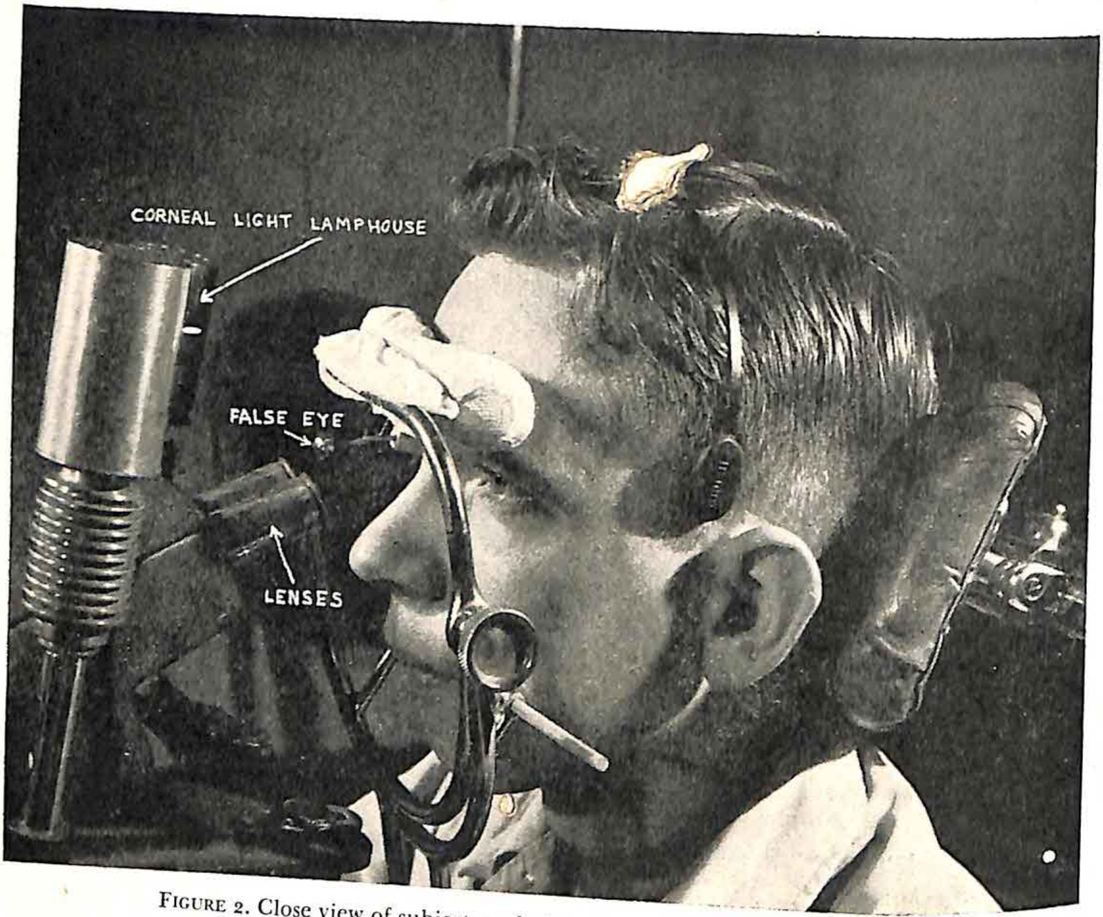


FIGURE 2. Close view of subject ready for making eye movement records.

it must be admitted that the bite stick did provide a very firm support and eliminated completely any rotating or lateral movements of the head. Furthermore, the subject did not change the bite once he had closed it on the stick at the beginning of the test run.

Several test runs of subjects looking at motion pictures were made using only the bite stick with the subject pressing his forehead against the forehead rest. It was found that the subject still tended to move his head slightly forward and backward, rotating in a vertical plane around the bite stick. This was eliminated by bringing up the headrest of the dental chair which applied pressure to either side of the head behind the ears by means of two small pads. These could be adjusted very snugly against the subject. Some subjects in the preliminary tests complained that the steel of the forehead rest was uncomfortably cold and hard, so it was covered with a thin layer of sponge rubber and a layer of gauze.

G. The Cueing Light

In the camera there was an aperture several inches to one side of the photographic aperture in which a small flashlight bulb could expose a small area of the film through a 1/8-inch aperture. The purpose of this cueing light was to make an identifying mark on the film at the moment when a particular test was beginning. Thus the camera could be put into operation some time before significant eye-movement measures were to be taken. When the actual test situation occurred, the cueing light could b[illegible]hed to mark the spot on the film. Later, in the analysis of the film, an allowance could be made for a standard number of frames between the cueing light and the camera aperture so that the first eye-record scene could be accurately identified. Eye-movement records were to be taken when the subjects viewed certain scenes of a teaching film. It would be necessary to mark a definite start frame on the test film at the beginning of each scene. At the same time it was necessary to flash the cue light in the eye camera so that in the eye-record film it would be possible to identify the particular frame where the subject began looking at the test scene in the teaching film. It would not be reliable to do this manually, i.e., to have the experimenter close the switch for the cueing light when the first frame of the motion picture came on the screen, because the reaction time of the experimenter could make a considerable error in locating the start frame when the motion picture would be going by on the screen at the rate of 24 frames per second.

For this reason, a device was constructed on the motion-picture projec-

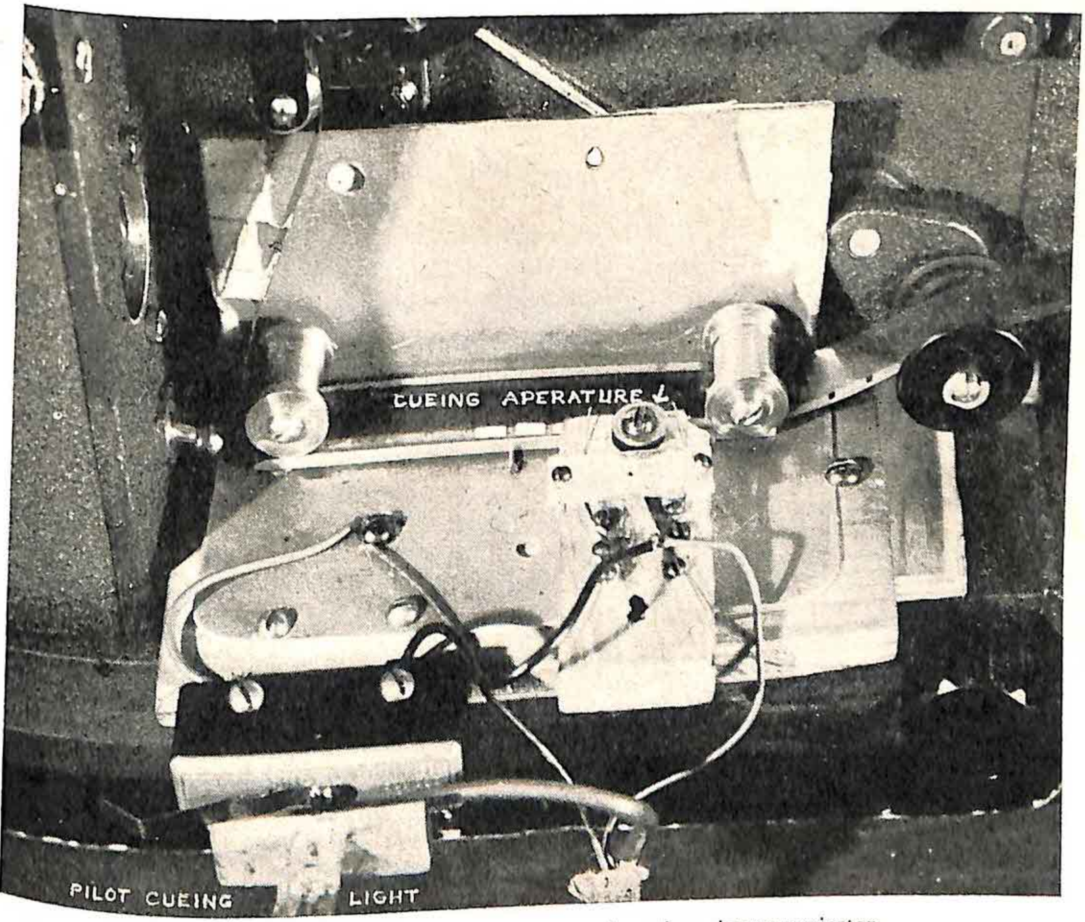


FIGURE 3. Detail of cueing device on sound motion picture projector.

tor so that the teaching film itself could activate the cueing light on the eye camera. It was decided that this cueing light should be activated and the cue mark put on the film exactly 40 frames before the start frame of each test scene. Forty frames were accurately measured on the bench and the frame for the activating device marked with a red grease pencil. At this point before each test scene three small pieces of tin foil were attached to the sprocket edge of the film by means of film cement. Three were considered essential in case one failed to activate the cueing light. These three marked frames passed through the sound projector in ⅛ of a second. A small pilot light was wired into the same circuit with the cueing light and attached to the motion-picture projector so that it would flash every time the cueing light flashed in the eye camera. In this way, the experimenter was certain that the cueing device had operated and that the flash frame had been put at the beginning of every subject's eye-camera record.

H. Registration Dots

It was necessary as usual in all eye-camera records to obtain several standard fixations at the beginning of each eye-camera record which serve as registration points for the subsequent analysis of the eye-camera record. This is customarily obtained in reading by first presenting the subject with a plain white card containing four black dots, one in each corner of the card and a short distance in from the edge. The subject is asked to look at these dots in a definite sequence, usually upper left, upper right, lower right, and lower left. These fixations appear in the photographic record of the eye movements and make it possible for the experimenter to estimate closely where in the copy the reader was fixating at the time he made any particular fixation dot.

[illegible]perimentation on motion pictu[illegible], it is even more important to have the same registration dots in order to locate the fixations in two dimensions and to determine the necessary enlargement of the eye-camera film when it is analyzed by projecting it on the screen. Therefore, a special strip of film was photographed on a title board, developed, and attached to the beginning of the test film. In this scene the registration dots appeared one by one in the usual sequence, starting at the upper left and ending at the lower left. The dots were placed some distance in from the edge of the frame so that imaginary lines connecting them would enclose three-fourths the area of the frame. Thus, the first thing that the experimental subject saw on the screen was these four dots. Before the test film was started, part of the instructions to each subject was a warning that these dots would appear and to fixate on each one as it appeared.

I. Preliminary Tests of the Camera

It was evident from the foregoing that several weeks were spent in adapting the Brandt camera to use with motion pictures. References have been made to continual tests of the operation of the camera, sometimes employing cooperating individuals and sometimes even making actual eye-camera records. However, when all the above changes had been made, fairly complete eye-camera records were made from three volunteer subjects to make a thorough test of the operation of the machine and to set up the technique and procedure of the subsequent analysis of the eye-camera film records. Several hundred feet of 35-mm. film were exposed in the eye camera while the subjects were looking at various motion pictures. The resulting film was developed and analyzed frame by frame. Particular attention was paid to the clearness of the image and to the movement of the false eye dot which registered head movement of the subject. These preliminary tests showed that the cueing device was not dependable enough, and it was redesigned until it operated perfectly. Dr. Miles Tinker was invited to examine some of the eye records projected on the screen, and it was his opinion that the method was feasible. These tests also showed the experimenter a number of precautions to be taken, such as keeping the subject restrained in the eye-camera headrest no longer than was absolutely necessary, giving plenty of time to get accustomed to the two corneal lights before giving the test film, etc.

J. The Reliability and Validity of the Method

The method described above of measuring eye movements is fundamentally the same as that employed by Langford (4) and Buswell (2). Buswell tested the validity by having subjects look very accurately at a series of five dots placed at intervals of 1 inch in both the horizontal and vertical planes. Forty-seven subjects then plotted these known fixations. He found that the error was 1 per cent in 72 per cent of the cases, and less than 3 per cent in 91 per cent of the cases. Although Buswell did not compute the reliability statistically, it seems that the level of error is extremely low. Weymouth and others (8) have found that the central axis of the eyeball and the line of regard approximate each other within 22 minutes or less. In addition, Dodge (3) has shown that the corneal reflection method is a reliable index of the position of the eyeball.

V. THE MAKING OF EYE-MOVEMENT RECORDS IN THE LABORATORY

The typical procedure followed in the eye-camera laboratory with each subject is as follows:

Before the subject enters the room, the eye camera is loaded with sufficient film to run the complete test on that subject. The camera is opened, exposing the film mechanism, and a series of circular punch marks are put in the film to indicate the number of that particular eye-camera record. This is entered in the log book. A fresh, sterile bite stick is inserted in the eye camera headrest.

The entire teaching film is threaded into the 16-mm. sound projector. Special care is taken in threading the film through the cue-mark device to make sure that it activates the mechanism. The cue-mark device is placed 40 frames ahead of the actual start frame of each of the sequences to be tested. In order to make sure that the lower loop below the picture gate of the projector is threaded to exactly the right size, two punch marks have been placed in the leader of the film exactly 40 frames apart. The film is therefore threaded so that one of these punch marks is located exactly in the picture-projection gate and the other punch mark is centered in the cueing aperture.

When the subject enters the laboratory projection room, his name is obtained, checked off on the group list, and entered in a record book, called the log book, opposite the number of the roll of film and the number of the "take" in the camera. It is explained to the subject that he is to be shown a motion picture and at the same time the eye camera is going to photograph the movements of his eyes as he watches the picture.

After the subject is put at ease, he is placed in the dental chair which is adjusted to support his back firmly. Then the elevator mechanism raises him exactly to the point where his brow fits into the forehead rest and he is able to bite the bite stick comfortably. Then the head clamp is brought up from behind and adjusted as tightly as possible without hurting the subject. It is explained to the subject that it is necessary to hold his head just as firmly as possible but that he must speak up if the head clamp is so tight that it hurts. The false eye is then placed over the subject's head.

At this point the lights of the room are turned out. One assistant takes the position in front of the screen and holds a small pen light exactly in the center of the screen area. The other assistant then proceeds to adjust the eye camera to the subject. He asks the subject to keep his eyes fixed on the pen light on the screen. Then he places another pen light over the ground-glass area so that its beam projects through the camera in a reverse direction from normal and casts a small spot of light near the subject's eye. Then the camera is adjusted by knobs to raise it and move it laterally until this light is centered exactly on the subject's left eye.

The corneal lights are now turned on. If their beams do not exactly hit the eyes, they are adjusted individually by loosening the rods that support them. The adjustment of the light into the left eye is only approximate but the right one is more carefully centered. The subject is then asked to look again at the pen light in the middle of the screen and finer adjustments are made on the camera. The corneal reflection is centered exactly at the crosshairs in the ground-gla[illegible]ture and then focused to as small and as [illegible] a dot as possible by means of the foc[illegible]g screw. As a final test, the subject is then asked to let his eyes roam around the screen and finally return to the pen light in the center. The action of the corneal reflection spot (later called the fixation dot) on the ground glass is noted to see that it moves only a normal distance to each side of the crosshairs.

The false eye is now adjusted so that its reflection dot appears in one corner of the ground glass. It is adjusted by manually bending the soft wire support on which it is attached to the head clamp.

The subject is now ready to view the picture. He is warned that the first thing he will see on the screen will be a series of four dots at the four corners of the screen in the following order: upper left, upper right, lower right, and lower left. One of the assistants illustrates by holding

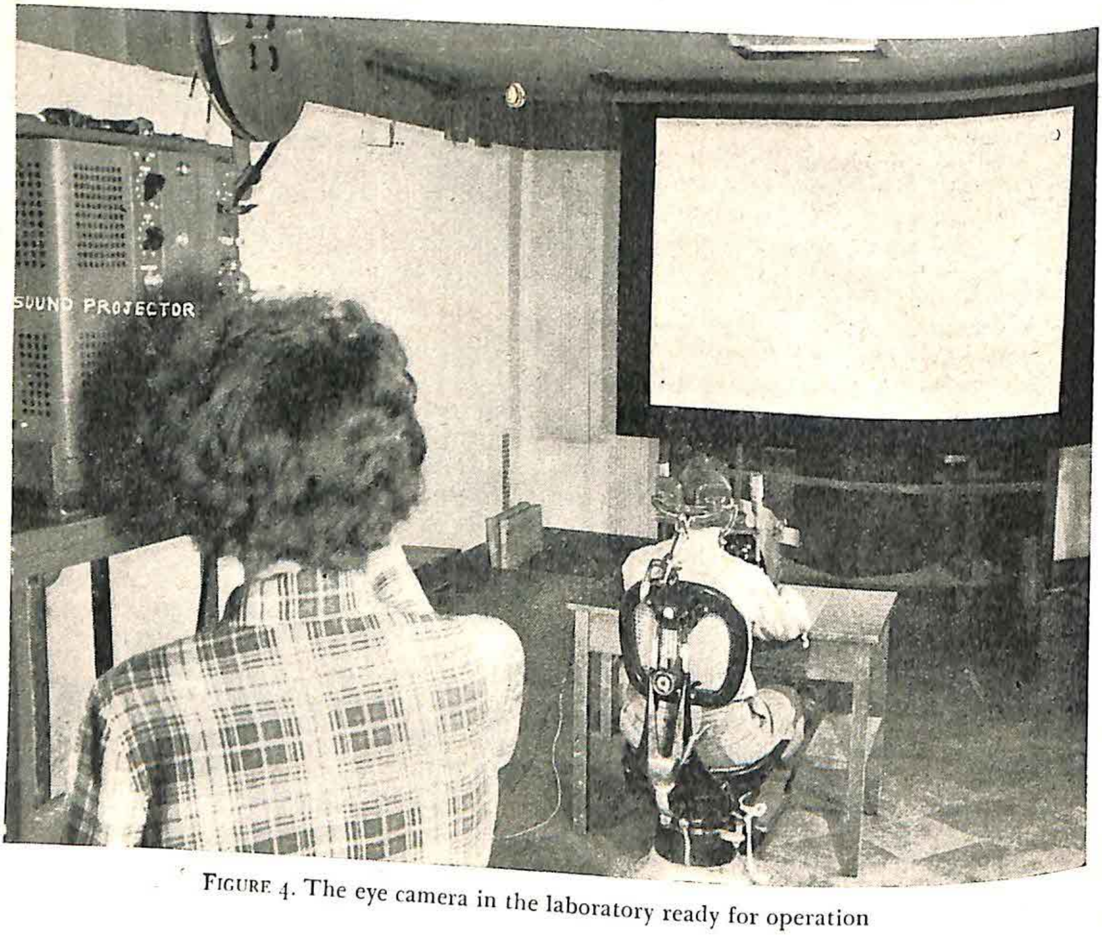


FIGURE 4. The eye camera in the laboratory ready for operation

the pen light in these corners of the screen before the film is started. The subject is instructed to look at these four dots carefully. After that he is to view the picture in an entirely normal fashion as he would any motion picture. The subject is asked to rest his hands comfortably on each side of the table or fold them in his lap, whichever he desires. It has been previously pointed out that there is plenty of leg room under the table for the subject to move his feet any way he wishes. During the running of the film and the eye camera, one assistant remains within reach of the eye camera and the subject although keeping out of the subject's range of vision. He watches the subject to see if there is any unusual reaction to the situation which might invalidate that subject's record.

The other assistant remains beside the motion-picture projector and controls both the projector and the operation of the eye camera.

The motion-picture projector is now started. The volume control of the sound-reproduction mechanism is turned as low as possible because the film leader containing the four orientation dots consists of silent-film stock. If this were allowed to go through the sound apparatus with the sound turned on, the perforations on the sound-track side of the film would cause a distracting noise known as "motor-boating." As soon as the four orientation dots have passed, the sound volume is turned up and the film is allowed to run then normally without interruption until the last test sequence has passed.

The operator of the projector is warned of the approach of a test sequence in each case by three small bits of white adhesive tape fastened to the sprocket-hole side of the film. The appearance of these going through the projector mechanism indicates that the cue mark preceding the start frame of the test sequence by forty frames is not more than five seconds from the cue-mark gate. At the appearance of these white bits of tape the operator turns on the eye camera. This allows a safe margin so that when the silver foil activating the cue-mark mechanism goes through the cue-mark gate of the projector, the eye camera is sure to be running smoothly. At the same time the cue-mark light is activated in the eye camera, the small neon light connected in the same circuit flashes beside the projector so that the operator knows that the cue-mark light in the camera actually has been turned on. Several seconds after the test sequence has ended, the operator turns off the eye camera. The camera is restarted at the beginning of each test sequence in the teaching film.

As soon as the final sequence has passed and all the eye records are obtained, the motion-picture projector is shut off. Then the eye-camera operator quickly opens the reflex gate on the camera which allows the eye dot and the false eye dot to appear on the ground glass, and checks them for position to make sure they have not moved off the field. The subject is released from the head clamp and bite stick and allowed to sit back in comfort to view the rest of the film. As soon as the film is finished, the lights in the room are turned on.

The footage meter on the eye camera is then checked to make sure there is enough footage for the next subject. (Usually three eye-camera records can be placed on a 200-foot roll.) Then the camera is turned on for a moment in order to allow some blank film between the two adjacent eye records. The 16-mm. projection film is rewound and rethreaded ready for the next subject. The film in the camera is punched with the next number of punch holes and entered in the log book as before.

VI. ANALYSIS OF THE EYE-MOVEMENT RECORDS

When the film is unloaded in the darkroom, the punch marks in the film are located by touch so they will not be lost with any clipping of the ends of the roll. The film is developed by hand in a Stineman spiral tank system using D-72 developer and developing the image 4 to 5 minutes. This time was previously determined by developing test strips from the camera for varying lengths of time to get the utmost density without too much fog.

After development, each subject's record is identified within each roll by its punch numbers, and the name of the subject is scratched into the film with a sharp instrument. Then the film is cut to separate each subject's eye records. The individual rolls are wound up on a flange, stripped off, and held together with white Scotch Tape on which the name of the subject is written with red grease pencil. The rolls are all wound head out. That is, the orientation frame is always on the outside of the roll. The rolls are then stored in 1,000-foot 35-mm. film cans which are sealed and temporarily placed in a film vault.

The next step is to locate on the 16-mm. print of the teaching film the exact beginning and ending of each of the test sequences. These had been roughly determined before the cueing material was affixed to the projection print, but now they are located exactly to the individual motion-picture frame. This is designated as the "start" frame. It is then necessary to locate the corresponding start frame in each eye-camera record. The first step in this process consists in determining the number of frames between the 16-mm. start frame and the cueing frame. This distance is determined for all test sequences and listed in the log book as "cue mark plus 40 frames," "cue mark plus 200." After these standard distances are entered, it is necessary to transpose this to the eye-camera film. Since the eye camera runs at 6 frames a second while the projector runs at 24 frames a second, the distance on the eye-camera record between the cue mark and the start frame is always one-quarter of the same distance on the projection print.

Thus, the projection print is used to determine the start frame and then, by counting backwards, the distance to the cue mark. This distance is divided by four and measured out on each eye-camera record starting with the cue-mark frame to locate the corresponding start frame on the eye-camera record.

The 16-mm. print is also used to determine the length of each test sequence measured in frames. This is again divided by four to determine how many frames of the eye-camera record after the start frame should be analyzed.

To facilitate the marking of the start frames on each eye-camera record, the different distances are laid out on a scale on a work bench. By using this, the start frames for each of the sequences in each ey[illegible]rd can be measured and marked qu[illegible] and accurately. The start frames are indicated on the film by outlining the frame and marking it with an X with a red grease pencil. Furthermore, the two frames preceding each of the start frames are marked with a red arrow pointing toward the tail of the film and indicating the number of the sequence whose start frame follows immediately after the arrow.

Two measures are read from these eye records: the amount of saccadic movement and the pause duration. On Dr. Tinker's advice, it was decided to call the amount of saccadic movement "interfixation movement." The term "sweep" cannot be applied because this refers to movement of the eye in reading in returning from the end of one line to the beginning of the next. The term "span" cannot be used here because this has been used specifically to refer to the number of words in reading comprehended in one fixation point. It was thought best, therefore, to use the term "interfixation movement" as applying to the straight-line distance between two fixations in viewing pictures of two dimensions.

It was decided to use the unit 1/24 of a second as the basic unit in measuring pause duration, simply because an empirical study of the eye records in a pilot experiment had indicated that there would be no pause durations shorter than this.

For convenience, two graph-paper data books are used in recording the interfixation range and the pause duration for each subject. Half of a double page

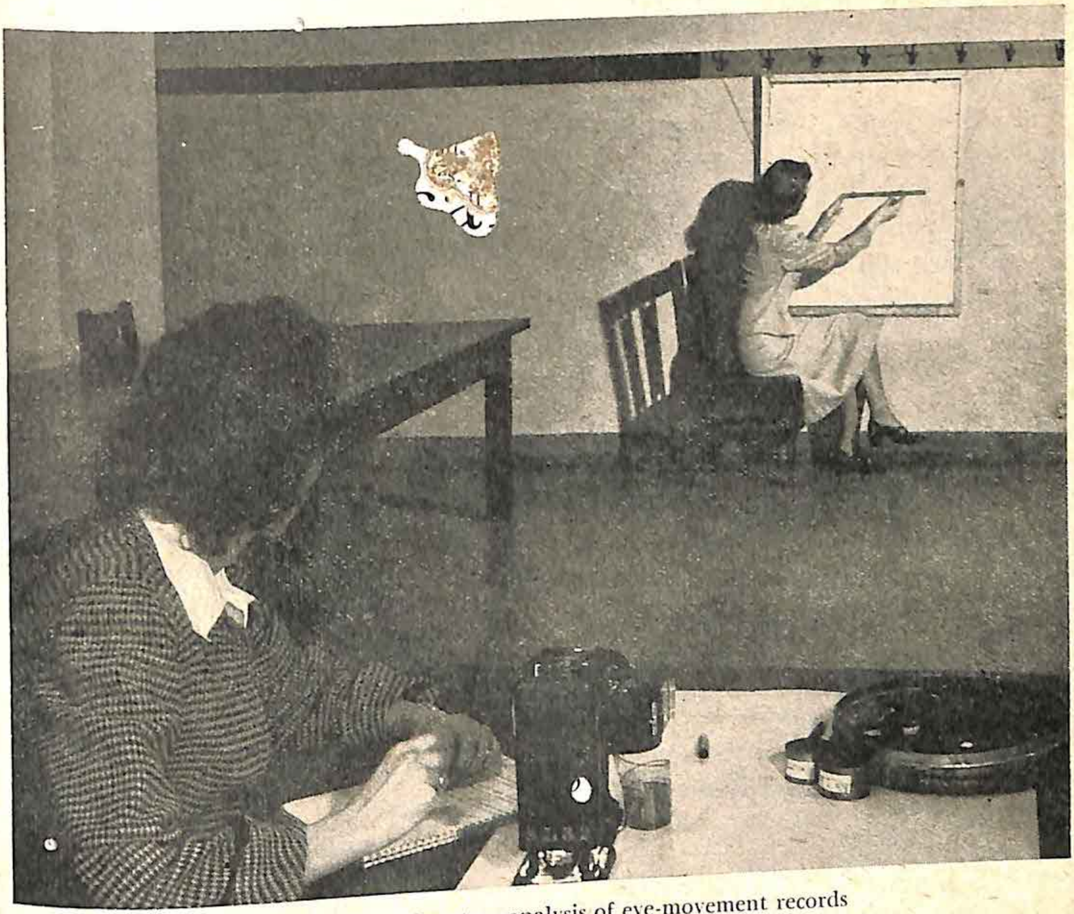


FIGURE 5. The laboratory analysis of eye-movement records

is assigned to each subject. Successive rows are labeled according to sequences. Vertical columns are numbered from one upwards to indicate frames out of the eye-camera film record.

The method used in reading the two measures of eye movement is to thread the eye-camera film into a 300-W. air-cooled filmstrip projector and project the film frame by frame on a fairly large screen. The projector is fitted with a single frame attachment so that only one frame of the film record is analyzed at a time, without any possibility of confusion. Because the eye camera moves the film horizontally through the eye-camera aperture, the eye-camera records have to be projected in a filmstrip projector which will similarly move the film from left to right through the filmstrip aperture. The distance from the projector to the screen, technically called the throw, is unimportant because there is no interest in the absolute movement of the eye which can only be measured in spherical angles; to be measured is the relative movement measured linearly on the plane of the screen without reference to the direction of movement. The relative eye-movement measures provide the data for the measurements of association between eye movement and other experimental factors.

It might be supposed that greater accuracy in measuring the interfixation movement might be obtained by moving the projector farther away from the screen since this would enlarge all the details on the screen. However, such great magnification is not useful because the factor limiting the accuracy of measurement is the sharpness of the image of the fixation dots on the film.

The screen consists of a piece of matte-white screen fabric fastened over a sheet of cork mounted on a solid frame. The cork is necessary to allow several pins to be stuck into the projection surface at one time. It is necessary to darken the room only partly because the powerful projector shining through the almost clear eye-record film produces a brilliant image on the screen.

Two assistants are necessary to operate this equipment. One assistant remains seated at one end of the room beside the filmstrip projector. Her part is to thread the individual eye-camera records into the filmstrip projector, advance the film to the start frame, and then advance it frame by frame as the data are collected. As each frame is projected, she tilts and moves the projector from side to side when necessary to superimpose the false eye registration dot over its previous position, thereby automatically correcting for the subject's head movement. After the record is run through, she rewinds it so that it is placed head out in the storage cans. It is also her responsibility to take charge of the data record book, entering the data as they are called out by the other assistant. She also assists the other worker in verifying the exact location of the eye dots on the screen.

The second assistant is seated at the other end of the room in front of the screen. When the start frame appears on the screen, the assistant at the screen, with the help of the assistant at the projector, locates the exact center of the false eye registration dot and puts a pin in the screen at that point. Then she locates the center of the fixation dot in that frame and puts a pin in that. The cle[illegible]he projector then advances the eye [illegible]ord one frame. The first step in reading the second frame, as in all subsequent frames, is to determine if the false eye dot is still centered on the pin marking it. If it is not, the projectionist turns and tilts the projector until the false eye dot is superimposed on the pin again. The location of the fixation dot in the second frame is made after the false eye dot has been realigned or immediately if this realignment is not necessary as happens in most cases. A pin is placed in the center of the fixation dot in the second frame. Then the distance between the fixation-dot pin established in Frame 1 and the fixation-dot pin established in Frame 2 is read aloud in millimeters. The assistant at the projector records this as the first interfixation range measurement for that particular scene for that subject. Then both assistants determine whether there are any other fixation points in that frame. If there are not, the pause duration is also entered as 4/24 of a second.

Occasionally more than one fixation dot appears in one frame, indicating that there was more than one fixation in the 4/24 of a second that that frame remained motionless in the eye-camera gate. It is easy to determine by the relative density of the two, three, or four dots whether the pause duration should be listed as 3/24, 2/24, or 1/24.

The collection of all our data on eye movement follows this procedure except for a few variations. One variation occurs when the fixation dot of the subject goes so far off to one side that it runs off the screen. This probably indicates that when the record was made, the subject was not even looking at the motion picture, but glanced away from the projection screen. In this case, the operators go back to the previous frame of the eye record and deliberately re-esta[illegible] the false eye dot a sufficient distance to one side or another so as to bring the eye dot back on the screen where it can be located. Then the measuring and the collecting of data proceed in the normal fashion.

Another interruption of the normal procedure is the occurrence of blinks by the subject. The evidence of this is merely the absence of any fixation dot whatsoever. This condition lasts approximately anywhere from 1 to 5 frames. The false eye dot is, of course, present in every frame, indicating that the absence of the fixation dot was not due to failure of the corneal light. Furthermore, the frame just before a blink always shows small light streaks radiating from the fixation dot. There are usually two light streaks extending in opposite directions from the fixation dot and somewhat resembling a blur caused in photographing a moving object. These peculiar "blink warnings" make it possible to differentiate a blink from the situation where the subject might look so far away to one side from the motion picture that the fixation dot would not be photographed at all.

In general, it is found that the false eye dot remains in much the same position. When it does move, however, it is characteristic of it to move down gradually from frame to frame and then suddenly to jump back up into its original position. The rigid control of the head movement of the subjects while the camera records are being made prevents any erratic head movements.

After all the eye records are measured, their accuracy is checked by a team of two other individuals who remeasure sample eye records chosen at random among all the subjects. One check showed that the original records were extremely accurate and dependable. The averages of the eye-movement measures determined for each student for each scene differed by only one per cent between the original analyses and the checking analyses.

At the conclusion of the data collecting, all the eye-camera film is stored again in fireproof cans, taped up and returned to the film vaults.

VII. SUMMARY

The Brandt camera (1) was modified to permit photographing eye movements in two dimensions by the corneal-reflection method while the subject is viewing a motion picture. The Brandt camera moves the film intermittently as in a motion-picture camera, but at the relatively slow rate of two frames per second. In the modified camera, this rate is increased to six frames per second. Synchronization of the eye-camera film with the motion picture being viewed is achieved by a cueing light. To permit analysis, the resulting film record is projected (by means of a film-strip projector) on a screen, where successive fixation points can be marked and interfixation movement can be measured linearly. Corrections for head movements are made at the same time by continually recentering the projected film image of a false eye worn by the subject.

This method of photographing eye movements of subjects viewing educational motion pictures is feasible and is recommended for use in further investigations in this field. The method makes possible the study of individual scenes and sequences of teaching films. Such study is needed to assure that the films have been planned and executed in a manner to achieve maximum teaching benefits at minimum production costs.

BIBLIOGRAPHY

1. Brandt, H. F. *The psychology of seeing.* New York: Philosophical Library, Inc., 1945.
2. Buswell, G. T. *How people look at pictures.* Chicago: The University of Chicago Press, 1935.
3. Dodge, R., and Cline, T. S. The angle velocity of eye movements. *Psychol. Rev.,* 1901, 8, 145-157.
4. Langford, R. C. Ocular behavior and the principles of pictorial balance. *J. gen. Psychol.,* 1936, 16, 307.
5. Society of Motion Picture Engineers. Recommended procedure and equipment specifications for educational 16 mm. projection. *J. Soc. Motion Picture Engineers,* 1941, 37, 4.
6. Tinker, M. A., and Frandsen, A. Evaluation of photographic measures of reading. *J. educ. Psychol.,* 1934, 25, 96.
7. Woodworth, R. S. *Experimental psychology.* New York: Henry Holt and Company, 1938.
8. Weymouth, F. W., Hines, D. C., Acres, L. H., Raaf, J. E., and Wheeler, M. C. Visual acuity within the area centralis and its relation to eye movements and fixation. *Amer. J. Ophthal.,* 1928, 11, 950-960.

VOLUME 66
NUMBER 8

WHOLE NO. 340
1952

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Factor Analysis of Some Reasoning Tests

By

HAROLD M. CORTER
North Carolina State College of
Agriculture and Engineering,
Raleigh, N.C.

Based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Pennsylvania State College, 1949

Accepted for publication March 5, 1951

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

TAB[illegible] CONTENTS

I. *The Experiment* 1

II. *Description of Tests* 4

III. *Summary of the Findings* 17

IV. *A Discussion of Test Composition* 22

V. *Clinical Validation of Factors* 25

VI. *Summary and Conclusions* 27

Bibliography 30

ACKNOWLEDGEMENTS

GRATEFUL acknowledgement is made to the many persons who have played a role in aiding the completion of this dissertation. Special thanks are due to Dr. Robert G. Bernreuter, Professor in the Department of Psychology, Pennsylvania State College, who served as chairman of my graduate committee. The other members of my committee, Dr. B. V. Moore, Head of the Department of Psychology; Dr. J. E. DeCamp, Professor in the Department of Psychology; Miss Mary Jane Wyland, Professor in the Department of Education; and Dr. W. U. Snyder, Associate Professor in the Department of Psychology, were helpful.

Dr. Nathan Jaspen contributed from his broad knowledge of statistics specific help in overcoming some of the hurdles involved in factor analysis. Fellow students helped by serving as guinea pigs for portions of the experiment.

Particular acknowledgement is made to the Junior class of the Bellefonte High School, Bellefonte, Pennsylvania, who served as the experimental group for this study. The staff members of this high school were very cooperative during the testing period and the author is greatly indebted to them. The two guidance counselors, namely Mr. Frank E. Hench and Miss Margaret Ferree, aided greatly.

H. M. C.

Raleigh, N. C.
August 8, 1950

CHAPTER I

TH[illegible]PERIMENT

A. INTRODUCTION

THE PRIME purpose of this study is to carry out an exploratory study in the area of reasoning, particularly in regard to various tests in the repertoire of the clinician. In terms of science in general, the procedure of factor analysis is one in which certain of the relational characteristics of reality are given some order and reference. In some ways it offers some hope that eventually psychologists will be able to speak in terms that other psychologists understand, and it may even some day result in a broad common ground through which all scientists can interchange ideas. Perhaps the most philosophically sophisticated of the people who have shown interest in factor analysis, Burt, defines the procedure in general as a means of providing parsimonious descriptions of psychological data (6, p. 249). Although Thurstone maintains that his search for "primary mental abilities" is not an attempt to resurrect an elementarism in psychology, in many ways he speaks about his factors as if they were such elements. In general the viewpoint is held by the present author that the mathematical analysis involved yields only data which have no inherent meaning and which must be interpreted by evidence from observation by trained people.

Two studies by Halstead (14) and Meili (22) present factors which, while somewhat abstruse, perhaps too highly abstracted, difficult to understand, and not completely identified, are ones that perhaps may eventually have more psychological usefulness than the highly objective test-material type factors now generally recognized. Highly objective factors in the traditional Thurstonian analysis have found some usefulness in the area of vocational guidance but have been somewhat sterile in terms of understanding either other individual behavior or its abnormalities. Descriptive terms which are satisfactory in terms of scientific knowledge should not only meet a criterion of objectivity in a parsimonious manner but should be able to contain or explain a broad range of human behavior if they are to be given the title of "primary mental abilities."

Prior to 1940 the results of factor studies have been subjected to a critical review by Wolfle (42). Since that time a well-organized summary of factor studies having relationships to aptitude and achievement has been undertaken by French (10). His analysis summarizes most of the studies made with young adults and adults. He makes an attempt to bring together various factors identified by different workers and renames them at times for cross identification. All in all, this book represents the contributions that factor analysts have made to our knowledge of psychology in digested form. The reader is referred there for further study.

B. PROBLEMS

One of the questions on which it was hoped that this study would throw some light is: Can reasoning be satisfactorily subdivided into separate factors?

Under this general question a number of possibilities might be true. Among these is the assumption that reasoning can be defined factorially in terms of the steps in the reasoning process. These are generally given in terms of problem-solving ability in which the successive steps are: first, recognition of the problem; second, bringing ideas to bear upon the solution of the problem; third, the weighing of various solutions to the problem; and finally, the testing of such solutions in practice. For many years logicians have separated reasoning into two components: deduction and induction. It would be of interest to know whether or not these a priori systems of analysis are meaningful to the extent that they can be recognized in terms of the statistical procedures used in factor analysis.

The design of the experiment makes it possible to make some empirical evaluations between typically objective test-centered factors as *"verbal," "spatial," "numerical,"* etc. against more generalized factors as found by Meili. The test battery contains five of Thurstone's tests which are supposedly relatively pure measures of factors found in his work. The tests also contain in the portion of the battery which is individually administered a number of tests which are similar to those used by Meili. Among these are the Comprehension test (No. 7 in the battery), the Naming test (No. 11), the Block Design test (No. 17), the Picture Titles test (No. 15).

Another question was concerned with the factorial content of tests not previously included in factor studies, most of which were tests which purported to measure reasoning, as, for example, the Arthur Stencil Design Test (No. 17), the Holsopple Concept Formation Test (No. 19), the Hanfmann-Kasanin Concept Formation Test (No. 20), and the concept-formation tests devised by the author. Another general question to be explored by the experiment was one concerning the relationship of concept-formation tests to the reasoning process in general. More than ten years ago Allport, according to Wolfle, questioned factor analysis because of its inability to meet a criterion of clinical usefulness or importance. Allport, in general, regarded factors at that time as statistical artifacts which have little or no relationship to real people. This complaint is still one which is felt to have some justification because of the rather limited usefulness of discovered factors. One aim of the present study is to bring the clinical and statistical approaches into somewhat closer alignment.

The largely rational, empirical approach used by clinicians perhaps needs some overhauling in terms of statistical strictness to prevent a meaningless verbosity; whereas, the highly objective statisticians need the application of empirical evidence from the clinic to make their manipulations have meaning. The viewpoint of the author is that the clinician needs to become a little more rigid in his use of language for the assessment of behavior, while the statistician needs to become more the psychologist and somewhat less the mathematician.

C. THE PRELIMINARY INVESTIGATION

A preliminary investigation utilizing a sample similar to the one to be used in the main investigation was undertaken in order to insure that the tests were of suitable difficulty, that they contained no errors or ambiguities, and that scoring methods and standards were suitable. A superficial examination of the results of this experiment was suggestive of fair possibilities in the group of tests involved.

D. THE POPULATION

The junior class of a high school in a town of approximately 7,000 population, formed the basis for the selection [illegible] experimental population. The Thurstone Primary Mental Abilities Test, Science Research Associates Edition, was administered to the entire junior class during one day of testing in the school. From the students present on the day of testing, a random selection was made and a list of 110 names compiled. From this group the 100 cases used in the experiment were selected in order. During the course of the experiment two students left school, one was not present on the final day of testing, so 103 names were used. Junior-class membership was divided into 46 per cent girls and 54 per cent boys. In the 100 cases used there were 57 boys and 43 girls.

This high school draws its population about equally from the town and townships that are adjacent to it. In the experimental group 43 pupils are transported by bus to the school. This is a rough way of estimating the number who are from rural areas.[1] The mean age as given by the pupils on the test blanks, where they were requested to record their ages to the nearest half year, was 17.46 years. This may seem slightly high for high-school juniors, until it is recalled that testing was conducted during the last two months of the school year. From the standpoint of achievement and maturity the group occupied a position that was midway between a junior- and senior-class group.

E. TEST ADMINISTRATION

Following the scoring of the paper and pencil tests and the selection of the sample, the high-school guidance advisors were requested to select four students, two boys and two girls, among the first fifty on the list whom they regarded as leaders in the class. These four persons selected were the first four to whom the individually administered tests were given. The purpose of this selection was to secure the cooperation of the remaining students to be later tested. Following the testing of these four persons, the remainder of the subjects was taken in convenient order. The place of testing was satisfactory, although upon occasion testing would be interrupted when movies, band practice, and other disturbing programs were held in the auditorium. All the testing, except for the final 20 subjects, was conducted by the author. With the final 20 subjects tested, the first 8 of the 16 individual tests were administered by a trained worker. With the first five subjects not tested by the author close supervision of testing procedures was made to insure standard administration.

In general, the test battery appeared to be appealing to the students as evidenced by the generally excellent rapport. How much of this was motivated by materials used in the test and how much was the result of students wanting to get out of regular classes was a question, but the over-all result was quite satisfactory. The average time of administration of the entire individual battery was approximately two hours at the beginning of the experiment. As the test material became more familiar, the examiner's efficiency rose to a point where, in most cases, the testing was completed in an hour and forty minutes, or two regular periods of the school session. In general, this period of time was sufficient to allow a pace that was unhurried but not wasteful of time.

---

[1] These data are based on attendance records on file in the school.

F. The Individual Test Battery

In general, it was felt that the individual tests should meet certain criteria in order to qualify as suitable tests. Owing to the fact that the examiner had to conduct the testing at rather irregular intervals, it was necessary to include only tests that were easily transportable. Ease of administration, objective scoring, and satisfactory scoring standards were other criteria used. Still another qualification of the test battery in general and of the individual tests themselves was that they should be administerable in a reasonably short length of time. In most cases, the latter was achieved to the point that no individual test in the battery consumed more than ten minutes. Tests which did not meet this criteria in their original form were shortened or modified. Another criterion was suitability in terms of difficulty. These criteria were, by and large, established in the preliminary study with a comparable group.

In order to insure comparable measures for each student and to minimize artifactual gains in performance due to gains in maturity, the testing altogether covered a period of only fifty-four days.

Chapter II

DESCRIPTION OF TESTS

The final test battery consisted of 21 tests, 5 of which were the subtests of the Science Research Associates version of the Primary Mental Abilities Test of Thurstone for group administration (Intermediate – Form AH [38]). The remaining 16 tests were all of an individual nature. They consisted, in general, of word tests, perceptual tests, and tests using various concrete objects. A more thorough description of the entire battery follows:

Test No. 1. SRA PMA Verbal test.
Test No. 2. SRA PMA Space test.
Test No. 3. SRA PMA Reasoning test.
Test No. 4. SRA PMA Number test.
Test No. 5. SRA PMA Word Fluency test.

These tests were included in the battery, largely for purposes of factor definition and to provide a means of comparing some traditional objective "Thurstonian" factors with Meili's factors.

An examination of the succeeding tests in the battery will reveal that all these tests except No. 4 had counterparts in the individually administered group. Test 1 and Test 8 were both vocabulary tests, and Test 10 had vocabulary elements. Test 2 had tasks felt to involve similar elements in No. 13, No. 17, No. 18, No. 19, and No. 20. Test 3 had, of course, most tests resembling it in the reasoning area. Test 5 was felt to be quite similar to No. 11, with some resemblances to No. 15.

Test No. 6. The Wechsler Form II Similarities test (40).

This test was included in the test battery to give a measure of verbal concept formation. Another reason was that it was felt that it would be valuable in establishing rapport, and readily accepted by the group. This proved to be the case.

Test No. 7. Wechsler Comprehension Form II (40).

This test was included in the battery because of its similarity with a series of practical questions mentioned in the Meili study and because it, too, was felt to be quite acceptable to most subjects and a good introductory test.

Test No. 8. A modification of the Stanford-Binet Vocabulary Test (27). The items from No. 14 to 38 inclusive of this test were presented. The instructions for the test were modified slightly and were as follows:

"I have here a series of hard words. Look at each word and tell me what it means. What does 'skill' mean?" Otherwise, administration followed usual Stanford-Binet standards.

This test was included in the battery to give added strength to any *"verbal"* factor and to compare the two different methods of test administration, group and individual, in terms of the factors in vocabulary tests. It was felt that the individually administered vocabulary test would require a combination of fluency and word knowledge.

Test No. 9. Judgment test. This test consisted of a series of ten cards, each with a ques[illegible] w[illegible] four suggested answers, which the subje[illegible] required to rank in order of which ans[illegible]peared to be best.

A preliminary selection of 20 items was made from Cardall's Test of Practical Judgment (7) that was felt to be suitable for students of a high-school level. From this group of 20 items, the ten which had the greatest weights according to Cardall's scoring standards were finally selected. These ten items were Nos. 10, 12, 15, 19, 20, 26, 37, 41, 44, and 47.

The instructions for this test were, "I have a series of cards that have questions on them. This test is to see whether or not you have good judgment. Look at each card and tell me which answer you think is best, which is second best, and which is third best. Try this one." After the first card the instruction was given, "You need only to tell me the correct letters in the correct order." (This was said if the subject gave the complete phrase rather than the letter.)

Scoring used Cardall's weights for these items.

This test was used in the battery because of its unknown factor composition and because some judgment measure that had face validity was felt necessary for definition and evaluation purposes.

It was felt, too, that it would provide a rather differentiating measure, and would serve to demonstrate the validity of the discrimination element in the reasoning process.

Test No. 10. Word Sorting test. This test was devised by the author to provide a measure of verbal concept formation. The test consists of a series of 12 cards, on each of which there are eight words, which can be divided into two groups according to a principle. In the first eight cards, the set which the subject has must alternately change from meaning of the words to some inherent characteristic such as the letters in the words, etc. A preliminary study was made with this test in an attempt to eliminate the possibility of having more than one correct response or method of sorting for each set of words. This attempt was successful, for no other means of classification were given by the experimental group which met the sorting requirements.

The directions for the word sorting test were as follows:

"I have here a series of cards with 8 words on each. On each card there are always two different kinds, or groups, or classes of words. What I want you to do is to tell me what two kinds there are on each card. Here is the first one. (Present demonstration card.) It will help you if you remember there are always four words of each kind on each card. It may be anything about the words; what the words mean; the way the words are written, or anything at all." At card No. 8 the instruction was repeated, "Remember it may be anything." The first card was explained if the subject failed to give the correct response or if he apparently had misunderstood the instructions. Similarly, when the first error was made, the criteria by which the words had to be sorted were reviewed for the subject. Most of these errors consisted of classifications with more or less than exactly four words in each group. Another question that was frequently necessary to ask was, "What is the other kind?" This occurred when the subject only classified one of the two groups on each card. The list of words was as follows:

1. FAT, hat, mat, pat, RAT, BAT, cat, SAT (Demonstration)
2. Betty, Paul, Ralph, Charles, Harriet, Helen, Harold, Florence
3. hay, height, obey, bite, sleigh, weigh, kite, right
4. beef, cow, hog, veal, deer, mutton, venison, sheep
5. race, runt, sack, sick, rope, rule, sole, sunk
6. John, Henry, Man, Woman, Person, Child, Genevieve, Susan
7. so, what, goes, no, or, duck, to, fill
8. stinking, sweet, foul, slimy, pretty, nasty, melodious, pleasant
9. wierd, slowly, silly, wiegh, funny, realy, Febuary, seldom
10. doves, children, men, oxen, muskrat, elephant, snake, eel
11. sad, lad, cad, fad, bad, mad, glad, dad
12. egg, look, bed, up, odd, silly, light, rang
13. editor, moon, lax, acting, betting, uncle, odd, many

In the scoring of these items, one credit was given for each group correctly categorized. Since each card had two groups of words alike in a certain respect, a total of 24 credits was possible. The subject was permitted to try the card as long as he appeared to be productive. Generally after a minute or so he would say, "I can't see anything there." In such case, he was instructed to try the next one. Testing was concluded after four successive item failures. In some cases a subject was able to name four of the words which belonged to one group and to give the principle by which they were grouped, but was unable to give the principle for the other four. In such cases, a credit of one was given.

On the sample item the two groups of words are words written in capital letters and words written in small letters. In the second item, four of the names are women's names and four are the names of men. The similarity in sound provides the basis for sorting of the third item, since there are four words which have an *ite* ending and four words which have an *ay* ending. In item four, four of the words are names of animals and the other four are the names of kinds of meat. In item five, four of the words begin with *r* and four begin with *s*. Proper nouns and common nouns or equivalent expressions thereof are the two groups used in item six. Four-letter words and two-letter words are the two groups in item seven. In item eight, four of the words have pleasant connotations while the other four are unpleasant. (This was mostly expressed by the subjects in terms of "nice" and "not nice.") In item nine, four of the words in the group are purposely misspelled; the other four are spelled correctly. In item ten, four of the words are singular and four are plural. Adjectives, or describing words, and nouns, or names of things, are the two groups in item eleven. Four of the words in item twelve contain double letters, while the other four have no letters repeated. In item thirteen, four of the words are begun with consonants and the other four with vowels.

This test was included in the battery to give a sorting test using verbal material that would be analogous to the Buttons test, the Opposites test, and to the various block-sorting tests used. It was felt to combine elements of verbal factors and the hypothesized reasoning factor or factors.

Test No. 11. Naming test. This test was another devised by the author in an attempt to measure a crude sort of concept formation with certain deductive elements, again using verbal material. It has been noted that early concept formation is apparently built around the abstractions made in the use of language, the assigning of material to various classes.

The instructions for the test were, "In this test, I am going to ask you to name all the things of a certain kind that you can think of in one minute. For example, I might ask you to name *white* things, and you would say 'snow,' 'clouds,' 'linen,' or 'paper.' Try this one. 1. Name as many red things as you can." The test was then continued with 2. round things, 3. sharp things, 4. cold things, 5. soft things, 6. elastic things, 7. things that there are generally five of in a group. During the testing, at the first pause in each group, the subjects were urged to continue by saying, "anything that is——will do."

One credit was given for things that are always or usually associated with the named characteristic. One-half credit was given for things that are occasionally or sometimes associated with the characteristic. The total score on this test was the total number of credits in all the categories mentioned. Examples of one-credit scores in the first category are fire-trucks, stop-lights, and finger-nail polish; half-credit items were such things as cars, houses, dresses, etc. In the second group, all things that have the general characteristic of roundness were given full credit, including items in which the shape was primarily spherical. The things given one credit were such things as rings, circles, etc. Half-credit answers here included such things as compacts, tables, etc. In the third group, both things that have sharp edges as axes, saws, etc. and things [illegible] were pointed as pencils and ice picks were g[illegible]redit. One-credit items are like those n[illegible]ed above, whereas half credit was given for such things as pieces of metal, etc. Cold things were credited in which the association is made of usually being cold as ice cream, snow, ice, etc., and half credit was given for such things as water, ground, or air. Soft things in the full-credit group included such things as pillows, sofas, mattresses, etc.; and in the half-credit group such things as food, chairs, etc. Elastic things included such things as rubber bands and balloons in the full-credit group, and such things as socks and other articles of clothing in the half-credit group. In things that there are generally five of in a group, such things as cents in a nickel, or men in a basketball team, or tires on a car (including the spare) were full-credit items. Such items as petals on flowers, or leaves in a group were given half credit, if the specific flower or leaf having these characteristics was not named.

This test was included in the battery to provide a measure of fluency, wealth of ideas, and deduction.

Test No. 12. Opposites (see Illustration 1). This test was devised by the author as an attempt to get at certain aspects of concept formation in a task which involves looking at drawings of various objects. The final group of drawings selected included four test cards plus one demonstration card. The cards were called circles, trees, squares, letters, and stamps. The task was to select from a series of five choices an object which was most unlike another object that was first presented on the card. In order to accomplish this, the first step in the process was to attempt to define the way in which the objects varied, and then to select the object which varied in the greatest number of ways. On the demonstration card two variables were given. A small black circle was followed by three numbered circles which were successively small and white, large and black, and large and white. Another item for additional practice was given on the demonstration card. On each of the four successive cards, which made up the actual test, there were four items which varied in terms of three variables per card, and the selection was made from five numbered items. On the trees card, the variables were the shape of trees (either rounded or triangular shaped), tall or short, and having light or black trunks. On the second card, squares, the variables were color (either red or white in the outer portion of the square), size of insert, and shape of insert (either a diamond or square in a rotated or unrotated position).

On the letters card, the variables were whether or not the letter was an *a* or a *b*, whether or not

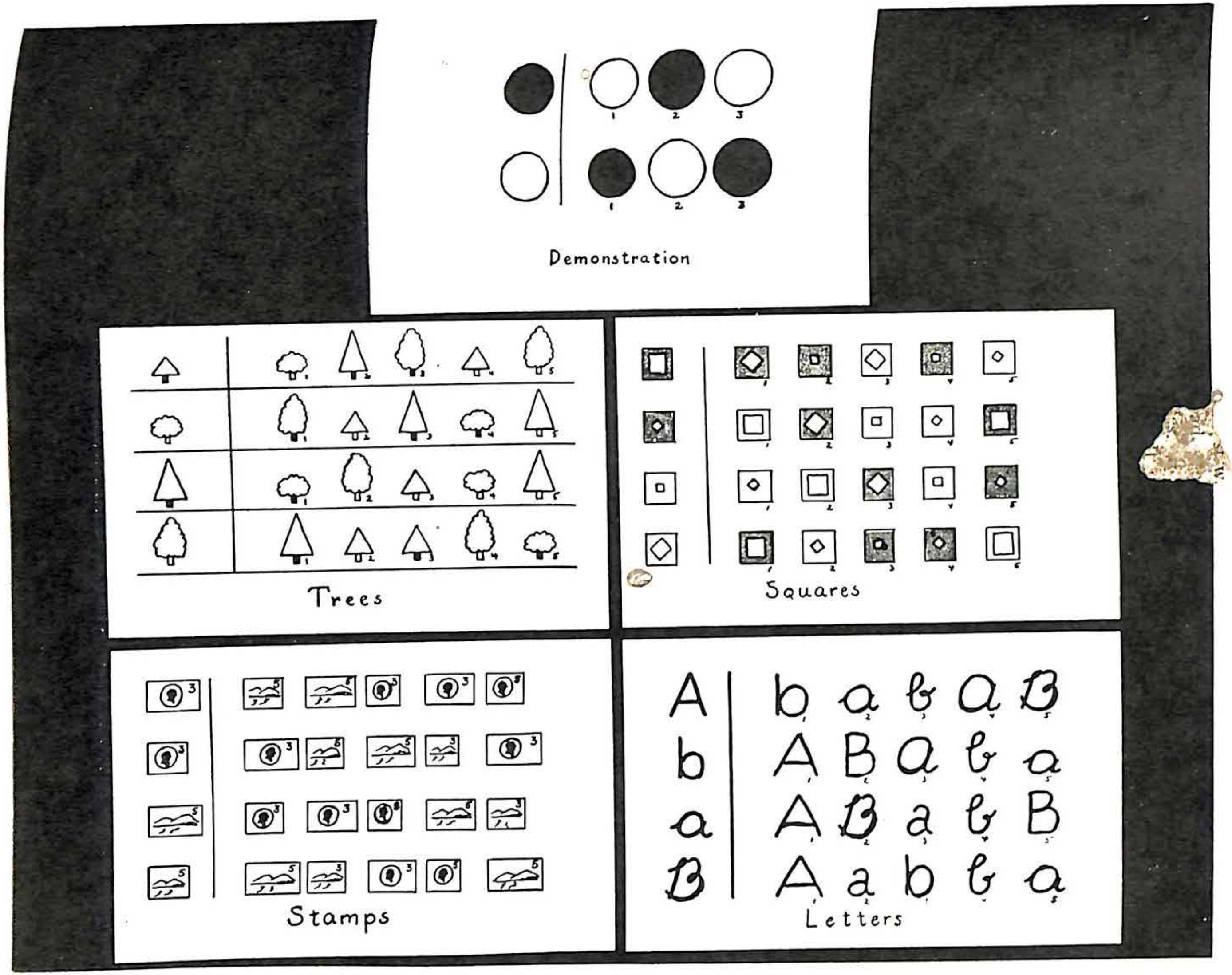


ILLUSTRATION 1. Opposites Test.

it was printed or written, and whether or not it was capitalized or a small letter.

On stamps, the variables consisted of length (long or short), the number on the stamp (5 or 3), and whether or not the stamp depicted a scene or a person.

Instructions were as follows: "Here is a series of cards that have various things on them (the first card was presented). Notice the first circle here is followed by three numbered circles that are different from it. Point to the circle that is *most unlike* the first one." If the correct response was given, the subject was reassured by saying, "Yes, that is right." Or, if he gave an incorrect response, the examiner said, "No, it is this one, because this one is not only larger, but it is also black," or made the necessary correction and added, "It is unlike in both size and color." The other item was tried by the subject in the same manner. The next card was then presented by saying, "Try this one." During testing, when a subject would begin to tell the ways in which the selected response was different from the test subject, he was informed, "Tell me the numbers of the ones on the cards that are most unlike the first one." Following the conclusion of this phase of testing, the cards were returned to the subject and he was asked, "What are all the things that you had to look for on this card in order to find the right answer?" The first time one of the variables was omitted, the examiner asked, "Was there anything else?" This question was only asked once. Occasionally a repetition of the question, "What are all the things you have to look for on this card in order to find the right answer," was made.

One credit was given for each item correctly answered in the first group, plus one credit for each variable correctly named during the second part of this test. The total score was the sum of the two types of scores.

This test was included in the battery as a concept-formation test that had the anomalous characteristic of finding differences in addition to finding similarities. Its factor composition was unknown, but was felt to revolve around reasoning and it was included as a perceptual task which would measure concept ability by means of recognition of difference.

Test No. 13. Gottschaldt Figures. This test was an adaptation for individual administration of the first 12 items on the SRA Research Edition of this test for group administration (31). These items were cut out of the Research Edition and pasted on 3 x 5 cards with the preliminary figure located to the left of the four succeeding figures in which it is to be found. For demonstration items the instruction page figures were used. These instructions were presented on three different cards which became successively more difficult. The items were then numbered for convenience in tabulating. Thurstone has described the characteristic that this measures as the ability to break perceptual gestalts (29).

The instructions for the test were as follows: "I have here a series of cards on which some designs are hidden in some more complicated designs. Look at this card (the first card in the demonstration series is presented). As you see, all of the designs here in this row have a square in them (pointing to the squares). I have made pencil marks so that they stand out more clearly. In this next group, designs No. 2 and No. 3 have a *V* like this in them (again pointing to the correct responses). This one has a *V* too, but it is not the same shape (pointing to No. 1), and this one (pointing to No. 4) has a *V* too, but it is not turned in the same direction, so that is not correct, either." (Card No. 2 is presented.) "This card is for practice. Which numbered designs contain the big plus sign like this?" (After the subject's response, correct any errors and give explanations.) "Try the next one on the card." (Again correcting errors.) The last demonstration card was then presented and the examiner said, "Here is a harder one like the ones you will have to do in the regular test. Try this one. Part of this test is to see how quickly you can do this kind of a problem, so you will have to work rapidly. Don't spend too much time on any of them. If you can't see it in number 1 rather quickly, go on to number 2, etc. You will have thirty seconds for each card, and this will generally be time for you to look each one over fairly carefully. You need only to tell me the numbers for each card. It may also help you to know that it may be in none of the designs, or it may be in all of them. There is no certain number on each card." (Present last demonstration card.) The subject then gave his response to the last demonstration card, any errors were corrected, any omissions pointed out and the subject was cautioned, "Sometimes you have to look pretty carefully." The first test card was then presented and at the conclusion of thirty seconds, the examiner suggested that the subject go on to the next card. On succeeding cards, however, he was not prompted to try the next card and generally chose his own pace, except in cases of extreme delay, where the time exceeded one minute.

The number of items correctly given according to the standards set forth in the examiner's manual for this test was used.

No time scores and no limitations of time were made except for the instructions. These instructions were designed not to provide a "speed" test but to prevent a subject dallying too long over each card. It was felt that, given sufficient time, all subjects could answer correctly, otherwise.

This test was included in the test battery because it was felt that it might have some connection with reasoning in the analytical ability required to perform satisfactorily and give some idea of the subject's ability to break a particular set.

Test No. 14. The McGill Picture Anomaly Series (17), Form N, the abridged version. This test was a series of picture anomalies or absurdities consisting of 21 items in the abridged form N.

Instructions for the test were simply, "Point to what is funny or out of place in these pictures." Additional instructions, "Use this brush so you don't make any marks on the paper," and "We'll only do the ones marked with a star or asterisk beside the number," were included.

This test was included in the battery because it was felt that it might play some role in delineating the role of rejection of hypotheses in the reasoning process, in which a large number of hypotheses are brought to bear upon the problem and the absurd ones are discarded. The recognition of error in general was felt to be of some importance in reasoning and its role needed definition.

Test No. 15. Picture Titles. This test suggested by one of the items in Meili's battery (22) consisted of a series of five pictures to which the subject was requested to give titles. There was also one demonstration picture. The pictures were from Wells and Ruesch's *Mental Examiner's Handbook* (41). They are *Bus Passengers,* by Isaac Soyer; *Morning on the Cape,* by Leon Kroll; *After the Tornado, Bahamas,* by Winslow Homer; *The Pioneer,* by Winslow Homer; *Rum Cay,* by Winslow Homer; and *The Nesters,* by Tom Lee. The *Bus Passengers* picture, by Soyer, was used as a demonstration card.

The instructions for the test were as follows: "I have here a series of pictures to which I want you to give titles. Give three titles for each one. (Present the first card.) For example, you might call this card, 'People,' 'At the End of the Day,' or perhaps 'Bus Riders.' Try this one." If he failed to give three responses, he was urged by the words, "Give three names for each picture."

Two credits were given for a title which subsumed the basic characteristics of the picture. One-point credit was given for titles which consisted of mere description or enumeration. On the first card, satisfactory two-credit responses were such things as "Morning on the Farm," "Beginning of the Farmer's Day," etc.; one-credit responses were given for such things as "School Girl," "Ploughing," etc. On the next card, "After the Storm" and "Shipwrecked" were two-credit answers, while "Desert Island" and "Dead Man" were examples of one-credit answers. On the next card, "Clearing the Forest" and "Pioneer" were given two credits, while such titles as "Woodchopper" and "Lone Tree" were given one credit. On the next card, two-credit responses were such things as "Catching the Turtle" or "Race for the Sea." One-credit answers were such things as "The Turtle" or "The Black Man." On the last card titles as "Western Farm Scene," "Pioneer Farm," and "Farm Out West" were two-credit answers, while "Farm Girl" and "Ploughing" were one-credit answers.

This test was included to give a measure of perceptual fluency or richness of ideas, or ability to synthesize the elements of the situation by abstracting the essential meaning. It was also felt, because of the request to give three titles, to give a measure of some of the power elements of intelligence.

Test No. 16. Rorschach (Beck's Z score) (4, pp. 58-82). In this test three of the series of Rorschach's ink blots, cards number II, III, and IX were presented.

The instructions for this test were "I have here a series of three ink blots. What I want you to do is to tell me what they look like to you, what they might be, and what they remind you of. Tell me three things for each card." The cards were then presented with the additional instruction that "You may use any part of the card, or turn the card any way you like to give your answers."

Subjects were urged to produce three responses by saying, "Look at it a little longer, perhaps you may see something else," when they stopped short of the requested three responses. Following the presentation of the three cards, an inquiry for position and further definition of the organizational elements required by the Z score was made. Other Rorschach inquiry was omitted.

Scoring was according to the standards established by Beck. The three card scores were then combined to get a total Z score.

This test was included in the battery as a somewhat less concrete test of the abilities required in the preceding test. Too, it was felt that the ability to make combinations was of some importance in the reasoning process, and according to Beck has definite intellectual aspects. If this were so, it was felt that they would revolve around some sort of reasoning ability. It was felt that it might give some test of Meili's belief that factors could be applied to all of behavior and not just cognitive functions.

Test No. 17. Arthur Stencil Design Test (2). In this particular test not all of the cards were presented because of the time limitation. (The total testing time for a complete test is usually about 30 minutes.) The final series of cards selected consisted of numbers 1, 4, 5, 8, 13, 17, and the demonstration card.

Another change was from a time limit of four minutes, which Arthur uses, to one of two minutes in order to give a test of suitable difficulty for high-school students. The instructions for this test were as Arthur suggests.

Three credits were given for each design completed within 2 minutes. Additional time credits were given as follows:

1. 5^s-1	
4. 20^s-1	10^s-2
5. 20^s-1	10^s-2
8. 40^s-1	20^s-2
13. 40^s-1	20^s-2
17. 50^s-1	35^s-2

These were then summed for a total score.

Only the necessary stencils to complete the required designs were presented.

This test was included in the battery to give some idea of its factor composition since it purports to measure, according to the author, "logical reasoning." Another reason was that it appears to be a task rather similar to the Block Design test No. 18 and its factor composition could be compared with that of a test that has been previously subjected to factor analysis. Another reason for its inclusion in the battery was the fact that it was felt to be a good transitional test from the perceptual tests previously used and the tests involving concrete objects used subsequently.

Test No. 18. The Block Design test. This test is the Wechsler Form II Block Design Test (40).

The inclusion of this test in the battery was suggested by the variety of factor composition found by different investigators using a variety of this test.

Test No. 19. Holsopple Concept Formation Test (21). This test consists of 16 different blocks which vary in eight ways, each block containing four of the variables. The variables used are color (blue or red), shade (light or dark) or intensity, shape (square or octagonal), width, height, size of insert, color of insert (white or yellow), and shape of insert (triangular or circular). The blocks can be sorted into four groups in a total of some 28 different ways, using two principles. This test is an unpublished test still in an experimental stage. While it is presented here as No. 19, it was actually preceded by test No. 20, which served the purpose of acting as partial instruction for this test.

The procedures followed in the instructions were as follows:

"I have here some different kinds of blocks. These blocks like the others can be sorted into four piles using two different principles. However, with these blocks there are always exactly four in each group, and they can be sorted in a number of different ways. What I want you to do is to sort them into four groups any way that occurs to you." (If necessary, additional instructions were given.) Following the first spontaneous sort, the subject was asked to verbalize the basis of the sort. If he only mentioned one variable, or had some difficulty in verbalizing any variable, further questioning was conducted by asking, "What is the way in which all the blocks in this group are alike?" If the subject still failed to recognize the sorting principle, he was asked, "How does this group of blocks differ from this group?" Following this period, the subject was requested to "Now sort them in a different way from the way you sorted them the first time." Inquiry was conducted as with the first sort. Then the subject was prompted to sort them by saying, "As you see, these blocks can be sorted in many different ways. I am thinking of a way that is different from those you've already used. Go ahead, and start sorting the blocks to see if you can figure out the way that I have in mind. You might make some mistakes at first, but after awhile you will see the way I have in mind and be able to do them without error." The last sort was made according to an a priori judgment by the author concerning the difficulty of various sorting principles. These were in order (including the principles of height and width used in the test previously administered, No. 20):

1. height
2. width
3. color
4. shape of block
5. shade
6. shape of insert
7. color of insert
8. size of insert

While Nos. 1 and 2 are ordinarily among the more difficult of the sorting concepts, it was felt that since they were the principles in the previous test, they would be among the easiest sorting principles on this test. If the subject used these principles, he was given full credit. Other principles were used with less frequency. During the first two sorts it was possible to utilize only three of the principles. In some cases, however, four were used and in such case, as in the case where the three were used, the next two levels of difficulty were presented in the final sorting. In cases where the sorts used were among the more difficult ones, the easier ones which had not been utilized were employed for the final sorts, except that only one variable out of the first two, height and width, was ever employed.

Four credits were given for each of the first two sorts correctly done without error. Two credits were given for sorting combinations which employed two principles in two of the sorting spaces and other principles in the other two spaces. In such cases, the subject was reminded, "Look them all over to see if you have sorted the same way in each group." If he spontaneously rearranged the blocks correctly, a score of three was given. On the third sort, a score of four was given if the subject recognized the principle of sorting one block beyond the point where sorting had become determinate. Three credits were given two blocks beyond. Two credits were given beyond this point. If the principle was discovered (based upon the criterion of errorless placement in the remainder of the blocks) before the last four blocks, one credit was given. Sortings are possible in which the principle does not become determinate before the ninth block, and in such cases the one score was reserved for the last three blocks.

Scoring for the inquiry portion of the test was as follows: Four credits were given for adequate verbalization of both sorting principles. If it was necessary to ask the way in which all the blocks in the group were alike or to inquire as to the difference between groups, only one credit was given for each variable provided the subject was able to verbalize it adequately. No credit was given in cases where verbalizations were inadequate. The total score on the test was the sum of the sorting score and the verbalization score.

This test was included in the battery because of its unknown factor composition, and the fact that it purports to measure concept-formation ability. It is different from the next test (No. 20) in that it covers a much wider range of material and has a characteristic of flexibility lacking in the other. The testing procedure was very flexible and permitted a spontaneity not found elsewhere in the battery.

Test No. 20. Hanfmann-Kasanin Concept Formation Test (16). This test was modified by the author to save time and to provide a more objective scoring method. To make it a more practicable test, it was decided that the following directions and instructions would be given.

"I have here a group of blocks. There are four different kinds of blocks in the group. This one (picking up the yellow semi-circular block) is called a Bik. It goes here" (placing the block on the appropriate corner of a card divided into four spaces which was used as a sorting area). "It goes here where it says 'Bik.'" "This kind of block (picking up the large blue square block) is called a 'Lag.' It goes here" (placing the block in the Lag space). (Picking up the small white hexagon), "This kind is called 'Cev.' It goes here" (placing the block in the Cev space). "This kind of block (picking up the high narrow orange colored triangular shaped block) is called 'Mur'; it goes here" (placing it in the appropriate space). The examiner then continued, "Now I've shown you one of each of the four different kinds of blocks in the group. What I want you to do is to go ahead and put the rest of the blocks where you think they belong. Pick them up one at a time, be careful not to look at the bottom, and put them where you think they might go or belong. At first you may make some mistakes because you won't know the way the blocks are sorted. If you make a mistake, I will tell you where the block should go, and after a while you will be able to see the way the blocks are sorted and be able to do the rest without making a mistake. Go ahead now and try one where you think it might belong." Corrections were made as necessary, and the blocks were turned over to reassure the subject that the examiner was not making up a sorting scheme of his own at each mistake. This procedure was continued until all the blocks were sorted. The subject was then asked, "How do they go?" And if he hesitated, "What is the principle by which the blocks are sorted?" If he still was unable to see the sorting principle, he was finally asked how all the blocks in each particular group were alike. And if he still failed to perform satisfactorily, he was requested to point out the difference between groups by the instructions, "In what way are the Murs different from the Cevs?" If he still failed, the solution was explained to him. The latter phase of the testing was conducted largely as an introduction to the next test, No. 19.

The scoring scheme used in the test was a combination error-time score. The number of errors made was subtracted from 20 and the number of minutes to the nearest quarter was subtracted from five and multiplied by four. These two scores were then added for a total score.

This test was included in the battery because of its unknown factor composition, its fairly wide use in the clinical field as a test of concept formation, and the rather rigid requirements of its unique solution for its performance. In contrast with test No. 19, time elements were utilized in its scoring and the freedom of choice in the above test was not only unavailable here for the subject but was probably handicapping.

Test No. 21. Buttons (See Illustration 2). This test was devised by the author specifically to give a test that was similar to 10 and 12, using concrete objects rather than words or perceptual material as the test items. The test consisted of a series of ten cards (post card size) upon each of which there were a number of buttons which were all alike, except for one, in a certain way. An additional card used for demonstration contained six buttons which were identical except one. On the remaining cards the buttons were all different and the subject was required by the nature of the test to recognize the way in which all the buttons were alike and/or select the one that was different in a respect in which all the others were alike. On card No. 1 all buttons had the same diameter except one. On card No. 2 the buttons all had two holes in the center except one. On this button the holes were near the edge. On card No. 3 all buttons had two holes except one which had four holes. On card No. 4 all buttons were opaque or "colored" while the different button was transparent or colorless. On card No. 5 the buttons were all upside down and all possessed the common characteristic of having a projection for insertion of the thread that lay parallel to the face of the button except one

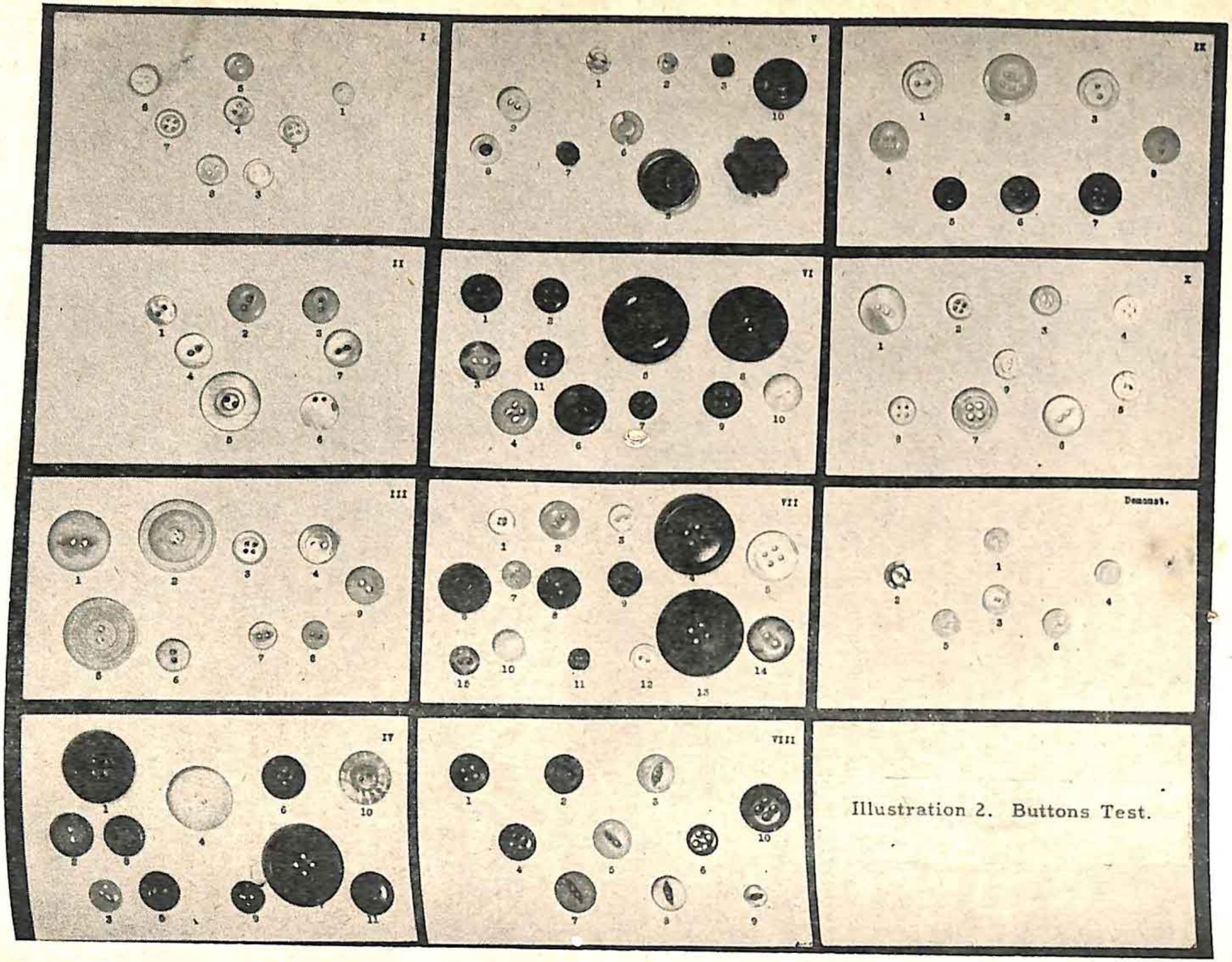


ILLUSTRATION 2. Buttons Test

which had the holes running perpendicularly to the face. On this card an alternate answer was also possible. On one of the buttons the material consisted of pearl and metal, making possible an alternate correct response of being made of two different kinds of material, whereas the others were all of one material. On card No. 6 all buttons were monochromatic except one which was two-colored. On card No. 7 all the buttons were round except one which was somewhat square shaped. On card No. 8 all buttons except one had grooves or indentations for the thread, probably to prevent the thread wearing out through friction. One button lacked this groove. On card No. 9 all buttons except one had some sort of rim extending around the outer periphery. On card No. 10 all buttons except one had a shiny appearance due to the fact that they are either simulated or real mother of pearl, whereas, the different button was somewhat dull, and made of a different material, perhaps palm nut.

The instructions for this test were as follows: "I have here some cards with various kinds of buttons on them. What I want you to do is to look at these buttons and tell me which button does not belong in the group. Look at this card (presenting demonstration card). All of these buttons except one are alike in a certain way." Point to the one which is different. After this was done say, "Yes, that is right, remember that all the buttons are alike except one, in a certain way. I want you to tell me which one is different." (Present first card.) "Which one is different here?" After the button was chosen by the subject, he was asked, "And how are all the other buttons alike?" The first time an incorrect response was made on a card, the error was shown to the subject by pointing out how the remainder of the buttons were not alike in respect to the way he had made his selection. He was then allowed to continue with the same card until he achieved the correct solution or was unable to see a solution, whichever the case might be. This procedure was followed for each additional card. However, only one such correction was made for each card.

One point was given for the correct selection of the "different" button and one credit was given for an adequate verbalization of the way in which all the other buttons were alike. In some cases where the subject was certain that he had selected the right button, but was unable to give an adequate explanation of the similarity among the others, he was given one credit. The total score was a summation of the scores of the individual cards.

This test was included in the battery to provide an analogous test to Tests No. 10 and 12 as mentioned above. In general, it was felt to be a test of concept ability using fairly common objects. On the basis of the preliminary testing, the test was found to be one which evoked considerable interest and left a feeling of general good will toward the test battery as a whole. It was a relatively easy test compared to some others in this battery and of genuine interest to practically all subjects.

A. STATISTICAL TREATMENT OF THE RESULTS

An investigation centering around the area of reasoning was undertaken. A preliminary selection of tests purportedly measuring this area was made. These tests were modified if necessary, for feasibility of administration to a high-school population. On the basis of this investigation tests were modified for the study. One hundred subjects were selected in the junior class of a medium-sized high school following administration of the Primary Mental Abilities test (SRA) to the entire class. The remaining sixteen tests were administered to the one-hundred subjects selected as an experimental group. Testing covered a period of approximately two months.

Pearsonian coefficients of correlation were calculated (see Table 1). Factoring was done using the group centroid method of Thurstone (32, pp. 149 ff.). The method outlined on these pages, including checks, etc. was followed closely. After the seventh factor was extracted, the residuals appeared to be too low to merit another extraction. In order to provide rotational freedom and insure the extraction of all possible meaningful factors an additional factor was extracted. This factor could be discarded later if necessary without causing too much computational adjustment and was felt to allow convergence of solution somewhat earlier than otherwise would have been the case. In general, in the author's opinion, in an exploratory study as this the extraction of all pos-

TABLE I

INTERCORRELATIONS OF TESTS

	M	σ	PMA V 1	PMA S 2	PMA R 3	PMA N 4	PMA W 5	Wech II Similarities 6	Wech II Comprehension 7	Binet Vocabulary 8	Judgment 9	Word Sorting 10	Concept Fluency 11	Opposites 12	Gottschaldt Figures 13	McGill Picture Anomalies 14	Picture Titles 15	Rorschach Z Beck 16	Arthur Stencil Design 17	Wech II Block Design 18	Holsopple C.F. 19	Hanfmann-Kasanin C.F. 20	Buttons 21
1	24.10	8.490																					
2	18.45	13.192	.281																				
3	14.18	6.160	.529	.279																			
4	20.11	8.212	.383	.136	.362																		
5	22.93	10.479	.467	.291	.427	.410																	
6	11.07	2.850	.441	.361	.306	.240	.343																
7	10.43	1.940	.173	.138	.149	−.024	.119	.210															
8	5.54	2.733	.632	.459	.473	.157	.250	.477	.288														
9	30.72	6.888	.193	.295	.290	−.009	.067	.014	.124	.269													
10	10.61	3.501	.480	.305	.450	.136	.230	.354	.232	.581	.372												
11	18.78	11.655	.174	.227	.170	.028	.253	.189	.205	.346	.096	.137											
12	17.63	4.379	.303	.348	.472	.057	.148	.328	.252	.435	.259	.345	.240										
13	6.72	4.398	.259	.207	.186	.040	.193	.191	.334	.240	.142	.238	.050	.200									
14	15.88	1.986	.268	.164	.105	−.083	.001	.205	.216	.377	.167	.183	.220	.156	.199								
15	14.85	3.254	.179	.277	.188	.031	.027	.230	.336	.317	.226	.223	.284	.130	.153	.342							
16	16.61	9.625	.148	.033	.074	−.052	−.018	−.048	.182	.148	.090	.191	.234	.090	.247	−.013	.240						
17	10.26	5.240	.250	.485	.296	.112	.228	.301	.139	.471	.291	.306	.242	.358	.444	.300	.223	.050					
18	23.92	7.953	.279	.445	.463	.148	.263	.264	.245	.306	.242	.229	.098	.324	.447	.155	.210	.104	.577				
19	20.83	3.023	.109	.291	.209	.125	.081	.261	.270	.195	.322	.296	.080	.186	.151	.067	.170	−.028	.228	.273			
20	19.63	6.640	.196	.246	.316	.087	.242	.182	.176	.280	.292	.276	.190	.284	.226	.002	.103	.015	.336	.262	.331		
21	13.66	2.899	.270	.360	.202	.131	.129	.275	.154	.442	.317	.369	.116	.284	.156	.201	.247	.068	.317	.257	.183	.113	

PE_r = 0–20 = .067
20 = .065
30 = .061
40 = .057
50 = .051
60 = .043
70 = .039

TABLE 2

CENTROID FACTOR LOADINGS (TABLE A)

	I	II	III	IV	V	VI	VII	VIII	h_s^2
1	645*	−434	−124	153	027	138	179	−112	7075
2	594	055	127	−184	−261	−143	−032	099	5052
3	629	−307	245	148	146	−173	−028	−152	6470
4	274	−421	217	240	−144	077	−076	101	3996
5	448	−306	200	307	−164	128	−144	−111	5049
6	544	−262	−085	−087	−205	212	−194	170	5329
7	413	232	−198	156	165	130	−170	126	3768
8	755	−173	−306	−196	−030	026	089	−137	7603
9	430	180	099	−244	253	−186	062	−058	3925
10	633	−172	−117	−183	307	034	097	070	5872
11	381	103	−200	085	−142	−092	−215	−276	3540
12	551	−040	040	−132	133	−180	−093	−118	3969
13	461	262	061	186	095	269	279	108	4904
14	350	139	−328	−091	−167	076	138	−179	3424
15	435	285	−354	097	−120	−144	−126	122	4711
16	194	160	−227	326	220	−229	164	106	3600
17	634	245	233	−146	−270	085	251	−102	6911
18	599	203	352	131	−096	−046	160	110	5901
19	401	160	155	−175	139	127	−257	219	3905
20	436	091	287	−082	189	126	−202	−174	4101
21	489	−031	−084	−241	−091	−130	133	173	3780

* Decimal points are omitted.

sible factors is justifiable. The unrotated factor matrix is shown in Table 2.

The method of extended vectors of Thurstone (32, pp. 225 ff.) was used for rotation because this method offers a rapidly convergent solution. Thurstone maintains that this particular method is suitable for batteries which show a positive manifold or positive correlations. It is not suitable for bipolar factors. Either an oblique or orthogonal solution is possible. Simple structure expressed in terms of minimizing as many loadings of tests on factors as possible was one goal of rotation. In general, the most important criterion used in the preliminary rotations was psychological meaning. For this purpose, the tests in the battery which had been found to have high loadings in certain factors were not allowed to influence rotations unduly, although it was felt that certain of the tests used in the battery for purposes of identification would have high weights in this analysis. In general, the tests showed good structure on the plots from about the third rotation and the remainder of the rotations was made to reduce angles, and obtain simple structure. One of the rotated factors was not

TABLE 3

COSINES OF ANGLES BETWEEN PLANES

	A	S	J	Pr	C	Rp	B	Fx
A	1.000							
S	−.012	.999						
J	.058	.042	1.000					
Pr	.195	.115	−.029	1.000				
C	.000	.003	−.130	−.009	1.000			
Rp	.030	.014	−.119	−.014	−.026	1.000		
B	−.014	−.031	.025	−.029	.042	.063	1.000	
Fx	−.032	.036	.033	.022	.002	.006	.010	.999

TABLE 4

CUMULATIVE TRANSFORMATION MATRIX

	A	S	J	Pr	C	Rp	B	Fx
I	397*	240	459	417	345	336	240	365
II	−792	347	−151	063	012	427	147	107
III	130	709	−053	012	−022	−368	440	−208
IV	229	004	−464	843	−143	057	−129	−194
V	−024	−053	311	011	203	159	−024	−870
VI	174	465	−304	−147	340	205	−672	−006
VII	−321	123	377	295	180	−690	−402	128
VIII	−108	−290	−466	−047	819	−152	311	−002

* Decimal points are omitted.

satisfactorily meaningful in terms of the criteria heretofore mentioned and consequently on final rotations it was necessary more or less to ignore this plane in order to achieve satisfactory solutions with the other factors. In general, simple structure was not complete in the battery as can be seen from an examination of Table 5. However, with some of the factors it is reasonably achieved. Thurstone points out that in experimental factor studies complete simple structure is rarely attained (32, p. 256). In general, the solution of the problem is orthogonal as can be seen from the angles in Tables 3. The transformation matrix is shown in Table 4.

TABLE 5

INTERPRETATION MATRIX (Λ)

Test	A	S	J	Pr	C	Rp	B	Fx
1	597*	034	383	407	191	012	−183	184
2	148	166	247	102	203	102	376	456
3	550	171	412	397	028	029	271	−012
4	556	079	−077	253	116	−124	082	078
5	601	227	006	371	−004	080	052	153
6	478	015	054	037	334	216	059	357
7	050	034	−018	243	278	471	042	020
8	344	−032	560	164	200	241	−049	396
9	−067	159	464	037	135	139	281	−008
10	290	000	497	121	394	150	051	017
11	149	004	162	196	−200	399	073	272
12	233	078	434	122	060	212	279	088
13	−031	369	085	405	384	116	−110	099
14	−042	012	260	110	013	213	−186	391
15	−072	−159	032	254	151	423	166	333
16	−114	−186	076	443	110	094	044	−097
17	006	516	342	223	173	040	126	504
18	084	436	158	425	266	014	315	246
19	090	218	039	−072	366	328	293	010
20	221	414	235	054	059	297	193	−066
21	071	−043	326	047	308	009	197	336

* Decimal points are omitted.

CHAPTER III

SUMMARY OF THE FINDINGS

A. FACTOR A

IT IS FELT that this factor can be best described as a learned activity factor that includes all those things that we ordinarily call *"academic"* in nature. Its primary expression in the test is through verbal material, but this fact is not surprising when we consider that most of our learning takes place through use of and by means of verbal processes. It involves the use of symbols and abstractions. In some respects it may represent the part of one's learning which is somewhat artificial or scholarly for use in school.

On factor A tests with substantial loadings are 1, 3, 4, 5, 6, 8, 10, 12, and 20 (see Table 5). This factor was most strongly identified in the Primary Mental Abilities portion of the battery. It was moderately strong in the adaptation of the Binet Vocabulary No. 8 and in the author's Word Sorting test No. 10. The highest weight is in the Word Fluency test of Thurstone and only slightly less strong in his Verbal test (Vocabulary). This at first suggests a *"Verbal"* factor until one attempts to explain the weight of the number and the reasoning test scores in Thurstone's tests. It would also seem to be a factor which is found almost exclusively in the type of activity required in a speeded paper and pencil test of the type used in this analysis, except for the fortunate circumstance that several of the individual tests have fairly substantial weights (No. 6, .478; No. 8, .344; No. 10, .290; No. 12, .333; and No. 20, .221).

This factor is quite similar to the "restrictive task" factor identified in Thurstone's analysis of perception (29) and is similar to the "central integrative" factor of Halstead (14) which he describes as the "well-organized habits of the individual." Halstead's belief that this factor has a high loading on so-called "psychometric intelligence" is thus justified if one confines oneself to paper and pencil tests and does not consider "psychometric" to include the individual tests in the repertoire of the clinician. Since Halstead derives his description from a wider range of material than was included in our test battery, he might have some justification for his somewhat broader factor name. In our battery the principle of parsimony and the lack of such range must limit our description to a so-called *"academic"* factor or learning factor. Certain of the tests which are quite similar in nature yield some support for the identification of this factor. For example, test No. 5, Word Fluency, in which the subject is required to name as many things as he can think of within a certain period of time that begin with the letter *s*, when compared with test No. 11, the author's naming test in which he is required to name red things, round things, etc. shows a rather substantial difference. In the case of the latter, the loading is only .149. According to some authorities both these tests should have loadings in and would primarily measure some sort of fluency factor. In terms of learning it is felt that initial letters are of considerably more fundamental nature than the enumeration of objects according to certain standards as required in test No. 11. The construction of dictionaries by alphabetical arrangement rather than by conceptual classifications is one illustration of this difference. Tests No. 19 and 20 show a somewhat less startling difference, but one which yields some slight corroborating evidence. The first test is a block-sorting test in which a number of variables are taken into consideration, while No 20 is a similar test with a single solution dependent upon the characteristics of height and width of the blocks. The way in which this test difference adds to the evidence is that one of the most usual methods by which objects are compared is by a criterion of size. The word "big" or the word "little" is felt to become a part of the young child's learned-response repertoire before such things as the other variables measured in test No. 19. In general, the tests which employ that which is new or unfamiliar or unusual have zero loadings on factor A. This is believed to be the reason why test No. 2 (the Space test) in the PMA series has relatively low loading, particularly when compared with the other subtests of this group test. Tests No. 7 and No. 9, both of which employ verbal material, are more difficult to explain, but it can be postulated that the requirements of these two tests are somewhat more discriminatory than is ordinarily needed in the course of one's educational history. Tests 13, 14, 15, and 16 all have slight negative weightings in this factor. In each case it can be argued that the material has little representation in terms of the psychological functions involved in learned activity. (It is of some interest to note that the Z score on the Rorschach has a weighting of −.114. The remainder of the tests, except for No. 20, were all relatively new tasks for the majority of subjects.

While this sort of a factor called *"academic"* in our analysis has been identified in other works as noted above, it is felt that the description herein given is well defined and identified by the tests in the battery. Subjecting this factor to our criteria for adequate factors suggests that this factor meets the standards satisfactorily. It should be of some interest to both clinicians and general psychologists as well as others interested in psychological functioning of human beings to be able to attain measures of the effects of one's learning experience. It might, if validated by further studies, be conceivable that if such a factor can be accurately identified, we could use it as one approach to study some of the shortcomings of our educational theory, and more particularly, our educational practice.

It is the opinion of the author that psychometrics must be broadened to measure other than the characteristics measured in the average group intelligence test. This may be difficult, but certainly the whole area of limitations of tests of particular types needs more airing. One has only to see people classified in terms of traditional tests as "borderline-feebleminded" functioning adequately in everyday social contacts to realize that tests sometimes do not tell the whole story. In our high-school population one of the surprises of the entire battery was to find that five persons had "quotients" of less than 70, the usual dividing line between feeblemindedness and normality. Many teachers would undoubtedly agree that they had "morons" in their classes, but they were probably employing the term somewhat less carefully than a psychologist should. One of the striking characteristics of some of these "borderline-feebleminded" subjects was that on some of the tests in the individual battery they achieved scores which were considerably closer to the mean than one would expect and in some cases were considerably above the average of the rest of their classmates. It is not felt that the author was fortunate to encounter a large group of idiot-savants among the hundred subjects used. It was felt that, instead, he was dealing primarily with academically handicapped people—in most cases probably poor readers. On the basis of social intelligence and other nonacademic activity, the "moron" group appeared to be relatively normal.

B. Factor S

This factor is likely most closely identical to some *"spatial"* factor in which the size, shape, or appearance of objects plays a predominant role. It is felt that this factor must be regarded as lacking entirely adequate psychological interpretation in terms of the particular tests employed.

Tests No. 18, 17, and 20 have rather substantial weights in this factor. Test No. 13 has a moderate weight and Tests 5 and 19 have slight weights. All these tests except No. 5 are tests in which concrete objects need to be manipulated. Only one test, test No. 5, is sufficiently different from the others to merit consideration, but this test has such a weak saturation (.227) that the rules of evidence rule out its having much weight. One rather remarkable weak test in this factor is the test employed in the PMA series to measure space. This suggests several alternatives, either the test does not measure space, or this factor is not a space factor, or it is not the kind of space factor that is measured by the other tests in this battery. The last alternative is believed to be the correct one. As mentioned above, the visual elements necessary to satisfactory performance on tests having substantial weights are primarily concerned with shape or size. In general, the very slight weight in No. 5 suggests that some people perform in this test by means of visual aids. This has some clinical validation in the behavior of subjects taking this particular test. One need not be a very keen observer to notice certain individuals performing on this test by use of sounds and one can actually see many students moving their lips and mouths to suggest that they are combining *s* words with various vowels to suggest to themselves words to be written. Similarly, one observes others looking around to get ideas from their immediate environment to aid them in the test. Similarly, one can easily observe a number of people in most groups taking this test who are "looking" but apparently not at anything in particular. Some even close their eyes, but by their expression, manner, and behavior suggest a visual activity. It is the last group which the author feels are achieving results in the test by the visual aid of seeing words beginning with *s* in their imagery. If the clinical analysis of this test performance can be validated, experimentally or otherwise, it would suggest a somewhat more psychologically oriented meaning for this factor, namely a *"visual imagery"* factor. As limned in this battery, however, we must on the evidence confine our interpretation to a spatial factor in which shape and size are of primary importance. Perhaps further exploration of this factor may divide people into some sort of dichotomy as that which the layman calls "concrete" thinkers and "abstract" thinkers, or it may lead to a division into groups such as auditory types versus visual types. Fortunately, the techniques of factor analysis do not make these groups antithetical but merely presuppose that they are uncorrelated, at least when found in an orthogonal solution. Perhaps of some importance in identifying this particular factor are slight negative weights in Test 15 (—.159) and in Test 16 (—.186). In the latter tests, visual imagery should not be a handicap unless it is a precise sort of imagery. It does lend some support, however, to the size and shape aspect of the interpretation of this factor because one has to allow himself liberty in distortion of both these characteristics to achieve adequate concepts for better scores.

C. Factor J

According to the author's analysis it would seem most likely that this factor can be most adequately described as *"judgment."* This judgment finds its expression in the weighing of hypotheses, evaluation of thought materials, criticism, and discrimination. It finds its expression in a wide variety of tasks in which *control* of thinking is of primary importance. The latter suggests a possible relationship between this factor and the concept of maturity or wisdom or the things one gains as he gets older and wiser, perhaps what we call experience.

This factor has high weights in Tests 8, 3, 9, 10, and 12, with moderate weights in 1, 17, and 21. This is obviously some sort of factor revolving around the sphere of reasoning. An examination of these tests is at first suggestive of some sort of inductive factor since test No. 3 is theoretically an induction test as are tests No. 10, 12, and 21. Similarly, one could present a fairly strong case for deduction in terms of its expression in Test 8 (the Binet Vocabulary test), No. 9, and No. 1, because some analysts have felt that in these tests persons bring a wealth of material to bear upon solutions by a deductive principle (see French [10] for example). The hypothesis of induction is not sustained by the weights that tests have that are readily interpreted as deductive in nature and vice versa. From the standpoint of an analysis using this sort of technique neither induction nor deduction would appear to be very clearly established.

It is felt that this name *"judgment"* serves to explain better the differences between tests which are somewhat similar in character, since Test 8, the Binet Vocabulary test, has a higher weight than Test 1, the PMA vocabulary test. One could argue that in the somewhat freer situation of the former that the subject needs to exercise considerably more discrimination, control, etc. than he needs to in Test 1, in which he merely has to choose among several alternatives instead of the mass of material coming to consciousness in the former case. Of the three concept-formation tests devised by the author, Buttons has the lowest weight, which suggests that the fewer limitations imposed upon the reasoning process in this test caused it to be somewhat weaker than Tests 10 and 12. The greater latitude in administration might also have played a role. Tests 17 and 18 might also be cited as additional evidence. In the first one a trial and error performance generally has much less effect than a similar performance on Test 18, where such a procedure is frequently quite valuable in achieving a correct solution—at least for some subjects.

D. Factor Pr

It is apparent that these tests are ones which center around two foci, namely speed of operation and fluency of ideation. Considering the two together, one can describe this factor as mental *"productiveness."*

On this factor, tests No. 1, 3, 13, 16, and 18 had high loadings, while tests No. 4, 5, 7, 15, and 17 had moderate loadings. All of the Thurstone tests are represented here except No. 2. Just as the amount of flow through a pipe is dependent upon the size of the pipe and the pressure, similarly mental productiveness is dependent upon speed and upon a number of ideas which come to one's mind that pertain to the problem. In tests in which the time is unlimited, the weight on this factor would be primarily dependent upon one's wealth of ideas. In a limited time, however, this factor would depend largely upon the degree of speed with which ideas reaching a conscious threshold are produced. Tests No. 1, 3, 18, 4, 5, and 17 are examples of tests in which the speed element is of paramount importance. Tests 16, 7, and 15 are examples of tests in which the richness of ideas causes heavy weights. Test No. 13, it will be recalled, was not a speed test and the subject was given unlimited time in which to perform. However, the instructions for this test were such that in essence, this test became a speed test. It is of interest that, allowed one's own pace, this factor still emerges. It is rather remarkable that some of the tests which are definitely speed tests do not emerge in this factor. One wonders, for example, why test No. 2, the PMA Space test, does not have a high weighting, because the score in this test is dependent upon the number of figures correctly identified as unreversed in the period of time allowed. It is of some interest to note that in this test a total of 19 subjects received zero scores. This includes, of course, subjects whose scores were negative in addition to those whose scores were actually zero. Since PMA standards were used in the scoring of this test, no attempt was made to evaluate scores below zero. In addition to these 19 subjects having zero scores, an additional 14 subjects had scores which were so low they might well be attributed to chance. This grouping on the test into two divisions, one of which scored significantly and a group which scored at a chance level, suggests a test which is perhaps too difficult or too crude in discriminatory power to be very useful. Part of the failure of this test to emerge with significant weight on this factor is perhaps due to some of these inherent weaknesses. Another test, in which speed was an element in scoring, that did not have weight on this factor was test No. 20. It was a rather common observation that among persons who worked rapidly on this test, the number of errors made appeared to be uncorrelated with the time, or perhaps was slightly negatively correlated. This test, however, does have more weight than its nearest likeness, No. 19, on this factor. Another test in which a time limit was used, not expressed in this factor is test No. 11, in which subjects were asked to name as many red things, etc. as they could think of in one minute. It was generally true that most subjects would exhaust their thoughts after thirty seconds on most of the items of this test. Yet, they were not given an unlimited time in which to perform as in tests No. 7, 15, and 16.[3] On the latter tests, more than one response is required except for some of the items in test No. 7. It was necessary for the subject in most cases to give these responses within a reasonable length of time. It was felt that in some cases subjects might conceivably have produced answers if they could have ruminated over the items, but the general pace of the tests suggested to them that they should not spend too long on individual items. This factor finds expression in a great variety of tasks, yet is uncorrelated with other factors. One might expect a factor in which speed is the most important element to be correlated positively with all other factors and lead to an oblique solution for the factor. A study of the angles (see Table 3) shows that while it has a slight positive correlation with our first two factors (*"academic"* and *"spatial"*), with some of the others it shows slight negative correlations. None of these correlations, however, is sufficiently high to suggest an oblique character for this factor.

The test battery in general was designed to permit emergence of possible speed factors. However, this was not the primary problem under consideration in the study, and, consequently, it did not emerge in our battery with as much clarity as in a study of this specific area by Davidson and Carroll (9). Their technique of using both speed and power scores for the same tests reveals this factor with more certainty. While our data do not approach this, it is felt that in general this factor is fairly well defined in the battery. As they point out, the implications of such a factor are many. Certainly we can no longer afford to make all tests speed tests or time-limit tests or power tests. In an adequate test battery, the common factor variance must not be seriously increased by such uniformity of testing procedure. In general, this factor seems to be relatively unchanged in shifting from one type of material to another. It is suggested that failure to take this factor into account has resulted in correlations among tests that are difficult to explain because they possess a spurious element, especially when speed is not a defined factor.

E. FACTOR C

With the exception of Test 13, these tests are ones which were supposedly measures of concept formation—the ability to grasp the essential similarity between discrete variables, to abstract them, and to generalize them. The idea of finding the rule or of reasoning from the specific to the general as induction, is apparently a similar way of saying the same thing. We have named this parameter the *"concept"* factor.

This factor was rather weakly identified in the test battery. Tests 6, 10, 13, 19, and 21 have moderate weights. Test No. 3 and test No. 12 are missing members of the group that was hypothesized as measuring this area. This ability seems to be, in our analysis, limited to types of tests in which choice is unlimited. This might serve to explain the lack of weight of test No. 3 and the failure of test No. 20 to match test No. 19 in loading. Test No. 13 is, in a crude sort of way, a test of finding similarities and might be on a superficial basis called a test of abstraction of a common element in spite of a mass of superfluous perceptual material. Test No. 12, which was hypothesized as a concept-formation test, has the characteristic of giving answers in terms of differences and this may be the reason why this test does not have much weight in this parameter. It is perhaps of some significance that test No. 11, a supposedly deductive test, had a negative weight of —.200.

---

[3] On a majority of the items in No. 7, however, two responses are necessary for maximum credit. On No. 15 and 16 three responses are required.

F. FACTOR Rp

This is operationally described in terms of willingness to respond where it was largely up to the individual to make his own rules of suitability in a deductive fashion. It might imply some suggestibility or overanxiousness to please the examiner, although this was not felt to be too important. It is also notable for a lack of a critical element in thinking. We call it *"responsiveness."*

This factor was neither clearly nor strongly identified in the battery, but the weights of the tests are rather strongly indicative of good psychological meaning for this factor. Tests No. 7, 11, and 15 are weighted heavily in this factor, while tests No. 6, 8, 12, 14, and 19 are significant. A review of these tests is suggestive of a factor which is perhaps only partly cognitive and perhaps partly temperamental in character. Recall of the directions and characteristics of Tests 19 and 20 will serve as some corroborating evidence for description of this factor as *"responsiveness"* since Test 19 allows considerably more freedom of choice and initiative on the part of S.

G. FACTOR B

This factor is too weakly identified in the battery to merit interpretation. It appears to be rather complex and is very slightly suggestive of some sort of analysis-synthesis dichotomy since its structure is somewhat bipolar. During the rotational process this factor caused a great deal of difficulty by its vagaries and it was not until it was ignored that rotation toward simple structure and orthogonality was easily achieved.

H. FACTOR Fx

This would appear to be a sort of ability to change one's mental set, or plasticity, or mental flexibility, or ability to try a new tack. It is felt that mental *"flexibility"* best subsumes this factor.

This factor is expressed in tests No. 2, 6, 8, 14, 15, 17, and 21. It was no surprise that test No. 17 emerged as having highest weight in this factor. After watching a great number of subjects attempt to reproduce the designs presented to them in this test in a reversed manner in which white served as a ground function rather than the figure, it was anticipated as some sort of inability to shift perceptually in terms of figure and ground relationships. Test No. 2, Space, as noted before, showed a rather remarkable dichotomy of distribution in terms of scores. From the author's own introspective analysis a great deal of shift is necessary to perform adequately on this test. The shift here, however, is not in terms of figure-ground relationships but is expressed in terms of rotational space. Somehow or other, the subject who performs well in this test transposes the figures among the answers to see whether or not they are congruent with the given figures. The great distress that a number of subjects taking this test always show is indicative of their inability to make this transfer. With this factor expressed in terms of these two tests alone one might assume some sort of a perceptual breakdown or analysis involved. Fortunately, however, it finds some expression in a variety of other tests, such as test No. 6, Similarities, and test No. 8, Binet Vocabulary, tests in which a considerable verbal element is involved. It is felt that the weight of test No. 6 is primarily due to the ability to eliminate superficial elements of similarity and to choose more fundamental (and better scoring analyses) on the early items of this test. On the later items, the inability to think in terms of similarities rather than differences is one important cause of failure. It is very frequently encountered in using a test of this particular type that dull subjects will persist in giving answers such as "they are not alike" and regard such an answer as entirely adequate. Such a performance is felt to be somewhat characteristic of people of average ability, too. Although they realize that some similarity might exist, they are so overwhelmed by the number of differences that come to mind that they are incapable of correct performance. The weight of test No. 8 is somewhat more difficult to explain. It is felt to have weight because of a difficulty in bringing to consciousness the essence of a definition with sufficient clarity to achieve a satisfactory response. It is generally recognized that multiple-choice vocabulary tests, as test No. 1, covering a similar range of difficulty, would likely result in higher scores than a test in which definition is required. The reader with experience in any kinds of tests will realize on the basis of his own experience that completion items are usually more difficult than choice items. It is felt that the reason for this is that possible answers are considerably more numerous in the free situation that exists in completion and the difficulty of settling on any one answer is consequently greater. Test No. 14 is a recognition of absurdity in pictures. The weighting of this test is suggestive of a similarity between this factor and that quality of individuals called a sense of humor. One of the characteristics of jokes is perhaps the sudden shift of set usually involved. Incongruity may be the essence of humor and perhaps the person who possesses a good sense of humor is one who is capable of recognizing absurdity. If this is so, it is notable that one dimension of intelligence or reasoning is necessary in order to get the point of a joke. Test No. 15, Picture Titles, has weight in this factor because of the necessity of the subject to think of some other way of describing the picture other than in the terms that he has previously employed. In general, one response was easy for all subjects; some people experienced difficulty in giving two responses; a considerable portion of the group was unable to give three responses, particularly for some pictures. The remaining test that had some weight was the Buttons test. In this test a different response is required for each set of cards and a new and different method of grouping must be employed. The persistence of memory of previous solutions was generally felt to be a handicap for a later performance. Those who gave satisfactory responses on some of the easier items on this test and were unable to recognize that it would be unlikely that the same principle would be used later gave less satisfactory responses. On the other hand, those who looked upon each item as a new challenge, with their minds relatively free of prejudices, apparently performed most adequately. It is of some interest to note that test No. 21 has greater weight than tests No. 10 and 12. The reason for this is that No. 10 and 12 place somewhat less emphasis and are somewhat less broad in the sense of the number of things to be taken into consideration than test No. 21. Test No. 16 has no weight in this factor because of the fact that shifts in set resulted, as often as not, in lowering the Z score. Most frequently the shift took place in terms of location of response rather than in terms of organizational ability involved in giving a new response having a high Z loading.

CHAPTER IV

A DISCUSSION OF TEST COMPOSITION

One aspect of the problem in this study was to discover the factor composition of various mental tests, particularly those which had not yet been subjected to factor analysis.

In general, the five Thurstone tests in the battery do not in our analysis emerge as "pure" measures. This is because, it is felt, the battery used did not have enough tests of a similar nature which could combine to allow for emergence of the Thurstone factors. However, they did show significant weights on some of the factors found in this analysis. It would thus appear that Thurstone's belief that tests rotated to simple structure in one test battery would retain their factor composition in other test batteries has not been upheld in this analysis. In the author's opinion, part of the failure of these tests was produced by limitations of the method of test administration, and the type and characteristics of the test itself.

Test No. 1 has significant weights on the *"academic"* factor, the *"judgment"* factor and the *"productiveness"* factor. In terms of the parameters used in this test it can be described as a measure of learned activity that requires discrimination and evaluation in its solution. Test No. 2, the Space test, in this battery was quite complex factorially and was expressed in the battery with moderate weight in the *"judgment"* area and with strong loading in mental *"flexibility."* It also had considerable strength in the unknown and unnamed Factor B and consequently might be assumed to have certain unknown loading. The Thurstone Reasoning test, No. 3, has high loadings in the *"academic"* factor, in the *"judgment"* factor, in the *"productiveness"* factor, and moderate weight in the unnamed factor. In this respect it would appear to be quite similar to test No. 1 in terms of what it measures. Test No. 4, Number, has a high loading on only one factor, namely the *"academic,"* and a moderate weight on the *"productiveness"* factor. Test No. 5, Word Fluency, measured *"academic"* ability, *"productiveness,"* and had slight weighting in the *"spatial"* factor of this study.

The Wechsler Similarities test was weighted with our *"academic"* factor, with the mental *"flexibility"* factor, and the *"concept"* factor. Test No. 7, Comprehension, was most heavily weighted in our *"responsiveness"* factor. It had moderate loading also in the *"concept"* factor. Test No. 8, the adaptation of the Binet Vocabulary, had expression in the *"academic"* factor, the *"judgment"* factor, and the *"flexibility"* factor. Test No. 9, Judgment, only found expression in the *"judgment"* area, although some of its variability was unaccounted for by its inclusion in the unnamed factor. Test No. 10, Word Sorting, was a fair measure of *"judgment,"* was among the strongest tests in the *"concept"* factor, and had modest loading in the *"academic"* factor. Test No. 11, Naming test, had its chief expression in the *"responsiveness"* factor. (It had a weighting of −.200 on the *"concept"* factor.) Test No. 12, Opposites test, is apparently a moderately good measure of *"judgment,"* with slight weighting in the *"academic," "responsiveness,"* and the unknown parameter. Test No. 13, Gottschaldt Figures, was a good measure of *"productiveness,"* and of *"concept"* ability. Test No. 14, Picture Anomalies, measured mental *"flexibility"* and to a lesser extent *"judgment."* Test No. 15, Picture Titles, measures *"productiveness"* to a slight extent and *"responsiveness"* and *"flexibility"* to a somewhat greater extent. Test No. 16, Beck's Rorschach Z score, appeared to be a relatively pure measure of *"productiveness."* Test No. 17, the Stencil Design test, turned out to be complex factorially, having important weights on the *"spatial"* factor, on the *"judgment,"* and on the *"flexibility"* factor, as well as a lesser weight on the *"productiveness"* factor. Test No. 18, Block Designs, measures *"spatial," "productiveness,"* and to a lesser extent *"concept"* ability and *"flexibility."* Some of its variance was unaccounted for by a moderate expression in Factor B. The Holsopple Concept Formation Test measures *"concept"* ability and *"responsiveness."* The Hanfmann-Kasanin test, No. 20, measures our *"spatial"* factor and to a lesser extent, *"responsiveness"* and *"judgment."* The Buttons test appears to be moderately good measure of *"judgment," "concept"* ability, and *"flexibility."*

In general, the predictions made concerning the factor composition of new tests were sustained. Few tests in the battery emerge as pure measures of any factor, except the Rorschach Z score, Picture Titles, and Judgment. This fact suggests two things: first, pure measures are difficult to construct, and second, our factors enter into a broad range of human behavior as abstracted in our tests. In general, they lose some value as differential diagnostic instruments because of this fact. The author makes no apology, however, for his inability to predict or limit his tests to a particular factor since the published and widely used tests of Thurstone fail to meet a similar criterion in this experiment.

If, as is maintained in this study, factors are primarily parsimonious descriptive phrases of "vectors of the mind," it is questionable whether the hope or attempt of some factor analysts to achieve a set of tests which are pure measures of particular factors will likely be achieved. Perhaps our approach to this dilemma needs modification toward the somewhat old-fashioned statistical technique of multiple regression.

A second alternative is to attempt to discover factors which have direct application for particular jobs that the psychologist must perform. This technique has been applied in a number of attempts to make up a suitable measuring and prediction device for vocational choice. In the author's opinion attempts to apply the findings of one area to other areas of interest to psychologists will prove either futile or inefficient. The failure of factorized tests to explain the neurotic, the psychotic, and organic behavior disturbances is perhaps due to the fact that they are not specifically designed for this purpose. It is not maintained that other tests of intelligence do a better job in this respect. It is sometimes distressing to the psychologist (but undoubtedly cheering to the patient!) to find that very gross destruction of cerebral tissues fails to change scores on tests. It is especially disconcerting when one observes the behavioral changes that take place in such cases and yet is unable adequately to define them in test terms. The failure to place one's finger upon the exact nature of changes and/or abnormalities of behavior reflects somewhat on our perspicacity as students of behavior.

Another solution to the general problem of prediction might conceivably lie in a direction of test design, particularly by paying somewhat closer attention to the administrative characteristics of tests themselves. For example, we might need to study the directions of tests in an experimental manner just as we study the test content. They may play a considerably more important role than we have realized. The findings of this experiment suggest that they aid in determining factor composition. Instructions to work rapidly have striking effects, for example. It would seem feasible to combine scores. For example, one test might give a part score toward a *"fluency"* factor, etc. and another test would yield part scores in these and other factors. By a method of statistical combinations the various weights could then be combined to give a more reliable measure of the parameters which have been found to have some applicability to a particular problem. It is conceivable that different combinations of the same scores might yield results useful in other areas. Again it is felt that the prescription will have to be specific. The hope of finding tests which measure only one factor and no other is believed to be somewhat more difficult than is commonly now realized, especially when one attempts to deal in the realm of the psychology of the individual. The great variety of approaches different individuals use in the same tests suggests an important difficulty. One only needs to discuss with ten different individuals how they go about deriving solutions to a letter-series test to discover that for some the solution is largely intuitive. For others it is a primary inductive task involving distance, while others see it has certain rhythmic characteristics and solve it that way. For perhaps one member of the group it might be a test of number series. Unless evidence is gleaned from a wide variety of sources, it is possible that pure measures of a certain factor might have vastly different implications from one individual to another. A variety of sources of weights toward a final factor score might minimize this difficulty.

It is difficult to see how some of the factors identified in this analysis can be purified. For example, it is difficult to imagine how the aspects of reasoning herein described can be abstracted from the contents of the tests themselves without disturbing the meaning of the factor. Reasoning in general must always be *about* certain things and is carried on *by means of* certain symbolic processes. It is difficult to see how one can construct a reasoning test that does not use verbal material, perceptual material, or concrete objects. Depending upon the type of material used, the test might then contain factors in addition to those involved in reasoning.

CHAPTER V

CLINICAL VALIDATION OF FACTORS

DURING testing of some individuals rather extensive notes were made, especially when the subject exhibited unusual test behavior as in cases of extreme difficulty or when he gave evidence of particular ability. This was done with two purposes in mind. First, it offered a means whereby some direction could be given in the initial rotations to eliminate unnecessary work. A forecast of eventual factor composition was made on the basis of the clinical evaluation. The number of rotations necessary (twelve) is indicative that this technique was not without some merit. The second reason why notes were kept was an attempt to give some means of validating the factors found in terms of behavior of individuals since it was felt that this might serve as corroborating data for the statistical procedures. If this sort of empirical validation could be made, it was felt that it would insure that the factors found would be psychologically meaningful, and perhaps more important, that they would be useful in the sense that they could account for deviations in performance. It is generally believed that scatter in test performance is primarily due to the differential abilities in factors which various individuals possess. It would seem that the converse of this—that individual performance could be accounted fcr by variance in factor ability—should also be true.

From the completed tabulation, the scores of six individuals were selected without knowledge of meaning of the scores by a psychologically unsophisticated person. These six subjects then served as a test group to see if there was merit in this belief.

The scores of these individuals on all 21 tests were transformed to z scores in order to make them comparable. These scores are listed in Table 6.

The first subject selected had three scores which differed by 1.5 standard deviations from his average z score of +.31. These are respectively Test 3 (—2 standard deviations), Test 7 (+3 standard deviations), and Test 13 (+1.5 standard deviation). Test 3 was a test which measured the *"academic"* factor, the *"productiveness"* factor, and the *"judgment"* factor. In this particular case, other scores measuring the *"academic"* factor are somewhat depressed as are the tests in the *"productiveness"* factor, except for those with unlimited time. Scores on the *"judgment"* factor are inconclusive here, but there is sufficient evidence on the basis of the other two factors to account for the majority of the variability. Test 7, according to our analysis, measures primarily *"responsiveness"* with a moderate weight in *"productiveness."* Since scores on the *"productiveness"* factor are somewhat depressed, the variability here cannot be accounted for by this factor. On the *"responsiveness"* factor, Tests 11 and 15 are good measures. The average of these two tests is somewhat above the average standard score for all tests and probably plays an important role in its high weight. However, it does not sufficiently account for the extreme variation here. Test 13 was a measure of our *"spatial"* factor, our *"productiveness"* factor, and of *"concept"* formation. Its elevation can be as-

TABLE

Raw Scores and *z* Scores

Tests

Subject	Type of Score	1	2	3	4	5	6	7	8	9	10
I	Raw Score	21	17	3	28	17	10	17	4	35	10
	z Score	− .37	−.11	−1.82	+ .96	− .57	− .38	+3.39	− .56	+ .62	− .17
II	Raw Score	16	24	9	13	0	8	9	2	40	8
	z Score	− .95	+.42	− .84	− .87	−2.19	−1.08	− .74	−1.30	+1.35	− .75
III	Raw Score	20	21	3	17	8	11	7	5	37	6
	z Score	− .48	+.19	−1.82	− .38	−1.43	− .03	−1.77	− .20	+ .91	−1.32
IV	Raw Score	13	23	9	8	19	11	10	3	35	4
	z Score	−1.31	+.35	− .84	−1.48	− .38	− .03	− .22	− .93	+ .62	−1.89
V	Raw Score	9	6	11	21	5	4	7	3	30	8
	z Score	−1.78	−.94	− .52	+ .11	−1.71	−2.48	−1.77	− .93	− .11	− .75
VI	Raw Score	33	9	4	24	13	8	9	3	24	8
	z Score	+1.05	−.72	−1.65	+ .47	− .95	−1.08	− .74	− .93	− .98	− .75

sumed to be due to the influence of the space element since other space tests are higher than average. However, concept formation scores are inconclusive.

In Subject 2, two scores of unusual character are present. Test 5 differs from the average *z* score by more than −1.5 standard deviation, while Test 9 is nearly +2 standard deviations above the average. Since in our analysis, Test 9 is a relatively pure measure of our *"judgment"* factor, its variability must be assumed to be due to presence to a high degree of such an ability. However, practically all of the other tests with high loadings on judgment are depressed. Therefore, we must regard this assumption as being disproven. Test 5, on the other hand, measures *"productiveness"* and the *"academic"* factor. In general, other tests measuring these factors are similarly depressed and it would seem to justify attributing the variability to these parameters.

Subject No. 3 has a score on Test 13 which is +2 standard deviations above the average. Since Test 13 measures *"concept"* formation, *"spatial"* and mental *"flexibility,"* its elevation would have to be attributed to these factors. Examining them in turn, it would appear that the spatial factor could account for a considerable portion of the variance, with some support from the mental *"flexibility"* factor. Tests of *"concept"* ability do not apparently contribute to the explanation for this anomalous score since they are ambiguous.

Subject No. 4 has one score on Test 15 at a point +2 standard deviations higher than the average of his other scores. Test 15, it will be recalled, measures *"responsiveness"* and *"flexibility."* Since other *"responsiveness"* tests are slightly depressed, the variability here must be attributed to the presence of ability to shift or the mental *"flexibility"* factor. The standard scores on other tests measuring this factor are considerably above average, so this would seem justifiable.

Subject No. 5 shows one score almost −2 standard deviations below his average. This is his Test 6 score. Test 6, it will be recalled, measures our *"academic"* factor, *"concept"* ability, and had a moderate weight in an unidentified factor. Other scores having weights in our *"academic"* factor are on the average considerably depressed, while tests having weight on the *"concept"* factor are slightly depressed. The combination of the two might well account for the variance.

Subject No. 6. The score on Test 1 is more than +1.5 standard deviations above average, while the score on Test No. 12 is more than −1.5 standard deviations below average. In general, these two tests measure different abilities, according to our analysis. Test 1 measures the *"academic"* factor, the *"productiveness"* factor, and to a lesser extent, the *"judgment"* factor. Support for the high score in Test 1 is not only lacking but in general is negative. The scores on other tests of a similar factor composition are in general depressed. Test 12, which measures judgment, is depressed as are practically all other tests measuring this variable.

In general, it would appear that on the basis of these particular cases, the evidence for clinical usefulness of our factors is slightly positive. In several cases, these factors fail to explain extreme deviations in scores. Out of the ten scores selected among the six individuals, six were readily explained, two were ambiguous, and two were contrary to what one would expect in view of other tests measuring similar factors. It is felt that

6

OF SELECTED SUBJECTS

11	12	13	14	15	16	17	18	19	20	21	M
19 + .02	18 + .08	15 +1.88	15 − .44	19 +1.28	34 +1.81	8 − .43	26 + .26	24 +1.05	19 − .10	14 + .12	+.31
9 − .84	12 −1.29	3 − .85	16 + .06	14 − .26	24 + .77	8 − .43	20 − .49	17 −1.27	19 − .10	9 −1.61	−.63
4 −1.27	18 + .08	15 +1.88	17 + .56	9 +1.80	13 − .38	13 + .52	30 + .76	23 + .72	12 −1.15	16 + .81	−.26
11 − .67	16 − .37	9 + .52	15 − .44	21 +1.89	32 +1.60	7 − .62	23 − .12	16 −1.60	8 −1.75	16 + .81	−.33
17 − .15	16 − .37	7 + .06	12 −1.95	13 − .57	27 +1.08	6 − .81	16 −1.00	23 + .72	27 +1.11	12 − .57	−.64
7 −1.01	7 −2.43	12 +1.20	16 + .06	17 + .66	17 + .04	3 −1.39	20 − .49	15 −1.03	10 −1.45	12 − .57	−.65

the ambiguity of this procedure merits further investigation, but it has not been proven that our factors are particularly useful in this respect. However, no means of comparison of the results with findings by other investigators are at present available.

CHAPTER VI

SUMMARY AND CONCLUSIONS

AN EXPLORATORY factor study in the general area of reasoning ability was undertaken to answer a number of questions, namely:

1. Can hypothesized reasoning factors be assigned to categories similar to (*a*) steps in a reasoning process, or (*b*) logical categories as induction, deduction, or (*c*) such descriptive terms as concept-formation ability, problem-solving ability, and analyzing and synthesizing ability?

2. What is the factor composition of certain tests which purport to measure reasoning or certain aspects of it?

Another aim of the study in general was to attempt to find factors which would have more psychological meaning in the sense that they would serve to explain a greater variety of human intellectual behavior, and would be more readily translatable into the descriptive concepts frequently used by clinicians. The position was taken that factors described in terms of test materials involved were inadequate and not sufficiently abstracted to be practical or useful except in limited circumstances. To this purpose, the statistical treatment of the data was always made subservient to meaning or usefulness. Two works by Meili and Halstead were especially recognized as excellent initial efforts in this direction.

A battery of tests was devised. They included the five subtests of the Science Research Associates version of the Primary Mental Abilities Test of Thurstone, which were administered in group form, and 16 individually administered tests. The five factors measured by the subtests of the former are: 1. Verbal, 2. Space, 3. Reasoning, 4. Number, 5. Word Fluency.

The remaining tests were:

6. Wechsler Form II, Similarities
7. Wechsler Form II, Comprehension
8. An abridged Stanford-Binet Vocabulary test
9. An abridged Test of Practical Judgment

adapted by the author from Cardall's test of the same name

10. A Word Sorting test devised by the author
11. A Naming test devised by the author
12. An Opposites test devised by the author
13. A Gottschaldt Figures test adapted for individual administration from the Thurstone Research Edition
14. The McGill Picture Anomalies Test, Form N, short test
15. A Picture Titles test devised by the author
16. Beck's *Z* score based on three Rorschach cards
17. An abridged version of the Arthur Stencil Design Test
18. The Wechsler Form II Block Design test
19. A modification of the Holsopple Concept Formation Test
20. A modification of the Hanfmann-Kasanin Concept Formation Test
21. A Buttons test devised by the author

The Primary Mental Abilities test, a group test, was administered to the members of the junior class in a medium-sized high school in central Pennsylvania. From this group, a random selection was made of 100 subjects who were given the remaining tests in the battery. Testing was conducted in the school in a suitable location during a period of approximately two months immediately succeeding the group tests.

From a correlation matrix, eight factors were extracted employing the group-centroid method as outlined by Thurstone. Rotation of these factors was made using the method of extended vectors as outlined by Thurstone.

The rotations, primarily determined by psychological meaning and usefulness, were made toward simple structure and orthogonal axes.

Seven factors were identified as a result of these rotations. A remaining factor was not capable of being interpreted. The first factor, A, was called an *"academic"* or *learned activity* factor. It found its expression in the test battery in vocabulary tests, number-computation tests, a word-fluency test, a reasoning test, and others. This factor was believed to be similar to a "restrictive task" factor of Thurstone and a "central integrative field factor" found in Halstead's work.

The second factor, S, was named a *"spatial"* factor having its chief expression in the ability to recognize size, shape, or congruence. There is suggestion but lack of sufficient evidence to identify this factor positively as some sort of visual-imagery factor. It was found in the battery in tests of Block Designs, Stencil Design, Gottschaldt Figures, and the Hanfmann-Kasanin Concept Formation Test.

The third factor, J, was called a *"judgment"* factor. It involves discrimination, ability to weigh and evaluate, to exercise control in one's thinking, and to function wisely as in a mature person. It had weight in the Reasoning test of Thurstone, the Binet Vocabulary test, the Practical Judgment test, the Word Sorting test, the Opposites test, the Stencil Design test, and the Buttons test.

A factor, P, called *"productiveness"* was primarily concerned with speed of output, although ideational richness played some role. It had weights in four of the five Thurstone tests, as well as the Gottschaldt Figures test, Beck's *Z* score, and Block Designs.

A factor, C, called *"concept"* ability was weakly identified. It apparently involves the ability to recognize essential similarities, to abstract and generalize, to think inductively. It appears to be somewhat restricted to tasks in which a considerable freedom of choice is allowed. It found expression in the battery in a test of Similarities, Comprehension, Word Sorting, Gottschaldt Figures, Holsopple Concept Formation Test, and the Buttons test.

Factor Rp was described as *"responsiveness."* It involved lack of criticism, some suggestion of willingness to please, and was felt to be partly temperamental and partly cognitive. Comprehension, Binet Vocabulary, Naming, Picture Titles, and the Holsopple test had appreciable weight in this parameter.

Factor B was the unidentified factor.

Factor Fx was labeled mental *"flexibility,"* or the ability to change one's mental set, or plasticity, or ability to shift. It was identified in Thurstone's Space test, Gottschaldt Figures, Similarities, Binet Vocabulary, Picture Anomalies, Picture Titles, and Buttons.

An attempt to validate the factors found in this study by an empirical technique based upon explanation of anomalous test scores was slightly suggestive of some clinical merit for our factors. However, the results were sufficiently ambiguous in this respect to state that the case has not been proven and further investigation is necessary along these lines.

A continued exploration by means of factor analysis in the general area of reasoning was recommended in order to clarify certain of the factors which were weakly expressed in the test battery used. This applies particularly to our *"concept formation"* factor and to the *"responsiveness"* factor. Further investigation is also needed to clarify the relative roles played by speed and/or apperceptive richness in our *"productiveness"* factor. Our *"judgment"* factor needs to be somewhat more clearly defined.

Recommendations were made that in future studies specific test batteries be designed for a particular area, as in clinical practice, because it is felt that other factors found using different approaches would have limited usefulness or application for such a specific situation. The suggestion was made that test constructors need to pay greater attention to the role of test instructions because of their probable effects upon factor composition.

It is somewhat difficult to see how some of the factors found in this analysis can be further "purified." For example, our judgment factor appears to enter into a very wide range of behavior. An adequate description of test performance would not only include the judgment aspects of the test, but would also have to include the types of materical used in the test itself. The elimination of nonrelevant factors would seem to be very difficult since some of our factors are ones which are apparently operative in a wide range of test performance. Perhaps a more pragmatic point of view would be to seek tests that have clinical usefulness or readily understandable psychological meaning.

As psychologists, we must be aware of the limitations of factor analysis as a procedure, as well as aware of its possibilities. The results of this study would suggest that there are many ways of classifying intellectual performance. Just as one can describe an elephant in terms of its trunk or legs or ears or tail, so we can describe only certain aspects of mental life. The only question of concern to psychologists should be, "What purpose does my description serve?" Test construction and factor analyses should be oriented to this end.

It is suggested that the parameters employed in this study be given empirical tests in the clinic to test their practical value.

BIBLIOGRAPHY

1. Army Air Forces Aviation Psychology Program Research Reports, Printed Classification Tests, Report No. 5. U. S. Government Printing Office, 1947. Edited by J. P. Guilford with the assistance of J. I. Lacey. Pp. 798-821, December 1942, July 1943, November 1943, September, 1944, Classification Battery; Pp. 142-155, Judgment Test; Pp. 180-190, Foresight and Planning Tests I; Pp. 180-190, Foresight and Planning Tests II; Pp. 215-225, Integration Tests; Pp. 333-339, Mechanical Tests; Pp. 261-268, Memory Tests I; Pp. 261-268, Memory Tests II; Pp. 408-418, Perceptual Battery I; Pp. 408-418, Perceptual Battery II; Pp. 113-122, Reasoning Tests.
2. Arthur, G. A non-verbal test of logical thinking. *J. consult. Psychol.* 1944, **8,** 33-34.
3. Arthur, G. *Stencil design test, I, II.* New York: Psychological Corporation, 1944.
4. Beck, S. J. *Rorschach's test. I. Basic processes.* New York: Grune & Stratton, 1944.
5. Blakey, R. I. A factor analysis of a non-verbal reasoning test. *Educ. psychol. Measmt,* 1941, **1,** 187-198.
6. Burt, Cyril. *The factors of the mind.* New York: Macmillan, 1941.
7. Cardall, Alfred J. *Test of practical judgment.* Chicago: Science Research Associates, 1942.
8. Corter, H. M. *A factor analysis of some individually administered reasoning tests.* Unpublished Ph.D. dissertation. Pennsylvania State College, 1949.
9. Davidson, William M., & Carroll, John B. Speed and level components in time limit scores: A factor analysis. *Educ. psychol. Measmt,* 1945, **5,** 411-427.
10. French, John W. *The factorial composition of aptitude and achievement tests* (Prepublication Draft). Lithographed. Educational Testing Service, January 1949.
11. Goodman, C. H. *A study of the Thurstone Primary Mental Abilities Tests.* Unpublished M.A. thesis. Pennsylvania State College, 1939.
12. Goodman, C. H. Factorial analysis of Thurstone's seven primary abilities. *Psychometrika,* 1943, **8,** 121-129.
13. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
14. Halstead, Ward C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
15. Halstead, Ward C. A power factor (P) in general intelligence: The effect of brain injuries. *J. Psychol.,* 1945, **20,** 57-64.
16. Hanfmann, E., & Kasanin, J. *Concept formation test.* Chicago: C. H. Stoelting Co., 1940.
17. Hebb, D. O. The McGill Picture Anomaly Series: Data on 100 unsophisticated adults. *Bull. Canad. Psychol. Assn.,* 1941, **1,** 47-49.
18. Hebb, D. O. The McGill Picture Anomaly Series, M and N. Orange Park, Fla.: Yerkes Laboratories, 1942.
19. Hebb, D. O., & Morton, N. W. The McGill Adult Comprehension examination: Verbal Situation and Picture Anomaly Series. *J. educ. Psychol.,* 1943, **34,** 16-25.
20. Heston, J. C. A factor analysis of some clinical performance tests. *J. appl. Psychol.,* 1943, **27,** 135-149.
21. Holsopple, J. Q. *Concept formation test* (unpublished).
22. Meili, R. L'analyse de l'intelligence (The analysis of intelligence). *Arch. Psychol., Genève,* 1946, **31,** 1-64.
23. Moffie, D. J. A non-verbal approach to the Thurstone Primary Mental Abilities. *J. gen. Psychol.,* 1942, **27,** 35-61.
24. Myers, C. S. A new analysis of intelligence: a critical notice. *Occup. Psychol.,* 1947, **21,** 17-23.
25. Pearson, K. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. Roy. Soc.* **CC.,** 1902, A. 1-66.
26. Rimoldi, H. J. A. Study of some factors related to intelligence. *Psychometrika,* 1948, **13,** 27-46.
27. Terman, L. M., & Merrill, M. A. *Measuring Intelligence.* Boston: Houghton Mifflin, 1937.
28. Thomson, Godfrey N. *A factorial analysis of human ability* (2nd rev. ed.). London: Univer. of London Press, 1946.
29. Thurstone, L. L. *A factorial study of perception.* Chicago: Univer. of Chicago Press, 1944.
30. Thurstone, L. L. Experimental study of simple structure. *Psychometrika,* 1940, **5,** 153-168.
31. Thurstone, L. L. *Gottschaldt figures.* Chicago: Science Research Associates, 1947.
32. Thurstone, L. L. *Multiple-factor analysis; a development and expansion of the vectors of the mind.* Chicago: Univer. of Chicago Press, 1947.
33. Thurstone, L. L. Primary mental abilities. *Psychometric Monog.,* 1938, No. 1, Chicago: Univer. of Chicago Press, 1938.
34. Thurstone, L. L. The effects of selection in factor analysis. *Psychometrika,* 1945, **10,** 165-198.

VOLUME 66
NUMBER 9

WHOLE NO. 341
1952

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Use of Vocational Interest Scales in Planning a Medical Career[1]

By

Edward K. Strong, Jr., Ph.D.
Professor of Applied Psychology, Stanford University

and

Anthony C. Tucker, Ph.D.
Colonel, Medical Service Corps, U. S. Army

---

[1] This research was conducted at Stanford University under Contract No. W-49-007-MD-483 with the Surgeon General, U.S. Army. The opinions expressed in this paper are the writers' and do not necessarily reflect those of the Department of the Army.

Accepted for early publication June 17, 1952.

Price $2.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

ACKNOWLEDGEMENTS

A TOTAL of 4005 physicians and 786 medical school seniors contributed to this research by giving their responses to items on one or more interest blanks. This very generous cooperation by busy members of the medical profession made this study possible and is gratefully acknowledged.

The American Boards of Internal Medicine, Surgery, Pathology, and Neurology and Psychiatry materially assisted in obtaining the necessary data by approving the project and our requests for the participation of their diplomates. The Bureau of Medical Economic Research of the American Medical Association made an important contribution by selecting a random sample of all physicians in the United States for use as a physicians-in-general reference group. The cooperation of the Association of American Medical Colleges in obtaining interest blanks from 786 medical school seniors is acknowledged.

The efforts of many individuals contributed to the conduct of this research, but space permits mentioning only a few. Special credit should go to Lt. Col. Fred J. Fielding, MC, Office of the Surgeon General, U.S. Army, for initially formulating the problem. He was largely instrumental in establishing the project and supplied valuable support at headquarters.

The following men served as graduate assistants on the project for the periods indicated: Robert E. Adamson, 2 years; Robert F. Fagot, 1 year; Ralph J. Garry, 1 year; Milton G. Holmen, 3 years; and John V. Zuckerman, 1 year. The Ph.D. dissertations of Holmen and Zuckerman resulted directly from their work on the project. These five men contributed much to the project.

A very large number of statistical computations were involved in this study and the efforts of Mrs. Lorena Nicholson in making sure that this work was efficiently and accurately performed are deeply appreciated. The large amount of secretarial work involved in obtaining data by mail was well done by Mrs. Bertha Bull and Mrs. Jean Biegenzahn.

Acknowledgement is made of the permission of Harper and Bros. and the *Journal of Applied Psychology* to reproduce certain tables.

EDWARD K. STRONG, JR.
ANTHONY C. TUCKER

CONTENTS

I. The Problem 1

A. Training Medical Specialists 1

B. Aptitudes and Abilities Versus Interests 1

C. How Permanent Are Interests? 2

D. Will Interest Scores Predict Behavior? 3

E. Questions Regarding a Medical Career 3

a. Should I Be a Physician? 4

b. Should I Specialize? 4

c. If I Am to Specialize, Which Specialty Should I Select? 4

d. What About Other Activities Within the Medical Profession? 5

II. The Development of the Interest Scales 6

A. The Primary Problem 6

B. Criterion and Reference Groups 6

C. Interest and Preference Blanks 7

D. Collection of Basic Data 7

E. The Interest Scales for Medical Specialists 8

F. Intercorrelations of Scales and Reliability 9

III. Differentiation of Groups 11

A. Standard Scores and Letter Ratings 11

B. Measures of Overlapping and Correlation 11

C. Differentiation with the Specialization Level Scale 12

D. Differentiation with the Four Specialist Scales 13

E. Differences Among the Four Specialist Groups on Other Occupational Scales 15

F. Results with Groups of Other Specialists 18

G. Results with Groups of Nonspecialists 20

IV. Practical Use of Individaul Scores 23

A. Reporting Medical Interests 23

B. Interest Patterns of Groups 25
C. Examples of Individual Records 27

V. SUMMARY AND CONCLUSIONS 29

TECHNICAL NOTES 33
A. Differentiation and Point of Reference 33
B. Development of the Medical Specialist Scales 37
C. The Specialization Level Scale (3) 45
D. Physician and Psychiatrist Scales on the Vocational Interest Blank .. 48
E. Development of the Medical Specialists Preference Blank 52
F. Cross Validation and Reliability 54

APPENDIX I 57

APPENDIX II 58

REFERENCES 61

CHAPTER I

THE PROBLEM

A. TRAINING MEDICAL SPECIALISTS

About two hundred graduates of medical schools enter the Army each year for internship training. Many of these interns later receive Regular Army commissions and take residency training in one of the medical specialties. A substantial number of Medical Corps officers have been certified by an American Board as diplomates in some specialty.

It is desirable that the Army should train the best two hundred of the approximately one thousand applicants—"best" meaning those who will profit most from further training and continue to serve satisfactorily in the Army. It is also desirable that these doctors shall be assigned to the type of training and work for which they are best fitted and in which they are most interested.

The research contract between the Surgeon General, U.S. Army, and Stanford University specified the "development of improved methods of guiding Army Medical Corps officers into specialty training. Emphasis will be placed on the development of measures of interest, motivation, and other personal factors, rather than on aptitudes or abilities." As it was not known whether medical specialists could be separated in this way, the project was restricted to the differentiation of four specialist groups, namely, internists, surgeons, pathologists, and psychiatrists. If such differentiation proved successful, further research was contemplated to differentiate other specialty groups.

Emphasis was placed on the development of measures of interest as indicative of what a person wants to do and will continue to enjoy doing. Research is being conducted under other auspices relative to aptitudes and abilities which indicate what a person is capable of doing. For example, the Association of American Medical Colleges and the Educational Testing Service are working on improving the methods of admission to medical school.

This project was conceived primarily from a guidance, not a selection, point of view. The Career Guidance Section of the Surgeon General's Office is continuously seeking better ways to carry out its functions. It was recognized that any improved methods developed for guiding Army Medical Corps officers into specialty training would be equally useful to young civilian doctors planning their professional careers. The methods developed by this research, therefore, are being made available to the medical profession.

B. APTITUDES AND ABILITIES VERSUS INTERESTS

In selection and guidance what a person *can* do is important. But what a person *wants* to do is equally important and far too often ignored. For example, men differ from women very little in intellectual capacities but they employ their abilities in different ways because they have different objectives in life—their interests are different. Graduate students in medicine differ very little in general intelligence and scholastic achievement from graduate students in business administration, but they have chosen entirely different careers. Table 1 consists of the distributions of scores of physicians, medical school seniors, and graduate students in business administration

TABLE 1

PERCENTAGE OBTAINING EACH RATING ON THE PHYSICIAN INTEREST SCALE

Rating	Score	Practicing Physicians	Seniors in Medical School	Graduate Students in Business Administration
A	45–70	71	73	1
B+	40–44	13	14	2
B	35–39	8	8	4
B–	30–34	5	4	10
C	0–29	3	1	83
	Total	100%	100%	100%

on the Physician Scale of the Vocational Interest Blank. The differences in their choices are most clearly indicated in their scores on the interest blank. Whereas 73 per cent of the medical students had an A rating, only 1 per cent of the business students had such a rating. The mean scores are, respectively, 51 and 22. (6)

The Physician Interest Scale is based on the responses of 500 physicians representative of all physicians in the United States. When a man takes this test, his responses are compared item by item with the typical responses of physicians; his score expresses how well his interests agree with those of physicians in general.

C. HOW PERMANENT ARE INTERESTS?

Does a man possess the same interests in later life that he had when he was a student? We are not here concerned with interest in some specific activity such as dancing or chess. We know that such interests are not particularly permanent. We are concerned with broad summaries of many specific interests, such as scientific, linguistic, and musical interests. Such interests are surprisingly permanent.

The permanence of scores on the original Physician Scale is given in Table 2 for intervals of one to twenty-two years. These data indicate that the physician interest score of a college man is likely to agree fairly closely with the scores he will obtain many years later.

When a college man is attempting to decide his future career, he is scored not on the Physician Scale alone, but on many other occupational scales. His choice should be made from those occupations on which he has the highest scores. The data in Table 3 indicate that a profile of 34 interest scores will agree very closely with a second profile secured up to 22 years later. From a guidance point of view the data in this table are more significant than those in Table 2 as they warrant the expectancy that the relative standing of occupational interest scores will be remarkably stable.

TABLE 2

PERMANENCE OF SCORES ON ORIGINAL PHYSICIAN INTEREST SCALE

Subject	Dates Tested	Interval in Years Between Test and Retest	Median Correlation
College Freshmen	1930–31	1	.85
College Seniors	1927–32	5	.78
College Seniors	1927–37	10	.77
College Freshmen	1930–49	19	.62
College Seniors	1927–49	22	.64

TABLE 3

PERMANENCE OF PROFILES OF INTEREST SCORES (7)

Subjects	Dates Tested	Interval in Years Between Test and Retest	Median Correlation
College Freshmen	1930–31	1	.88
College Seniors	1927–32	5	.84
College Freshmen	1930–39	9	.67
College Seniors	1927–37	10	.82
College Freshmen	1930–49	19	.72
College Seniors	1927–49	22	.76

D. Will Interest Scores Predict Behavior?

Reliability and permanence of test scores are important. But the most important measure of a test is the extent to which it predicts future behavior. Do college men who score high on the Physician Scale actually become physicians and do those who score low enter occupations other than medicine?

Among 670 Stanford University freshmen of 1930 and seniors and graduate students of 1927, whose occupational careers are known up to 1949, 108 became physicians and 562 did not. The distributions of their physician interest scores are given in Table 4. The mean physician interest score of those who became physicians is 47; the mean score of those who did not is 30, a very large difference. Among these 670 students there was a 53 per cent chance of becoming a physician with an A rating, 20 per cent with a B+ rating, 12 per cent with a B rating, 11 per cent with a B− rating, and only 2 per cent chance with a C rating. When all the extraneous factors such as ability, health, financial status, and family responsibilities are considered, it is amazing that any test could predict entry into medicine to the extent shown.

TABLE 4

Comparison of Physician Interest Scores of College Students Who Became Physicians with Those Who Did Not (8)

Physician Interest Rating	Students Who Became: Physicians		Non-Physicians		Chances in 100 of Becoming a Physician
	Number	Per Cent	Number	Per Cent	
A	70	65%	63	11%	53
B+	14	13%	56	10%	20
B	10	9%	73	13%	12
B−	9	8%	73	13%	11
C	5	5%	297	53%	2
Total	108	100%	562	100%	

Our data actually show a still better relationship than that indicated above between interest scores in college and occupation engaged in 20 years later. Men with relatively high physician scores, who did not actually enter medicine, definitely did tend to enter occupations closely related to medicine, such as dentistry, biological sciences, etc.

The Physician Scale is designed to predict the probability that a person would like or enjoy the practice of medicine. Since *practicing* medicine is somewhat different from *studying* medicine, there are men who like one but not the other. Several outstanding students in law school have had low lawyer-interest scores. Later on they entered other occupations, well foretold by their low lawyer scores. Because they were good students, they had enjoyed studying law but they discovered later on that they did not like the activities common to legal practice. A medical student with a low physician interest score may enjoy medical school, but he will probably not enjoy medical practice.

E. Questions Regarding a Medical Career

A guidance point of view implies an orientation in terms of the decisions to be made by a young man who is interested in medicine. There are four questions he must answer successively, namely:

a. Should I be a physician?
b. If so, should I specialize?
c. If so, what specialty?
d. What about other activities within the medical profession?

a. Should I Be a Physician?

This question does not merely mean, Does he have the characteristics of a physician? It must mean, Does he have the characteristics of a physician to a greater degree than he has the characteristics of an engineer, lawyer, accountant, and all the other occupations that college men are likely to enter? To answer this question, the college man should be scored on a battery of many occupational scales. He should seriously consider the pros and cons before entering an occupation on which he obtains a low score. Interest is not the only factor in vocational selection, but it is an important factor not lightly to be disregarded.

What is meant by the expression "characteristics of a physician"? The essence of being a physician consists primarily of those modes of behavior in which physicians differ from other men. The only men who should consider medicine as a career are college men or potential college men. We are concerned, therefore, with only the differences between physicians and men in the occupations that college men normally enter, not with all men in the United States. In developing a scale for physician interest, the likes and dislikes of physicians are contrasted with the corresponding responses of men in the occupations entered by the majority of college men. All items to which the two groups respond alike are ignored; the only items which are included in the scale are the items on which physicians differ from the men-in-general group.[1] A score on the Physician Scale indicates whether a man differs from men in general in the same manner that physicians differ from men in general or whether he differs in some other direction.

b. Should I Specialize?

Specialization generally requires three years of residency training and study preparatory to passing an examination. It also requires the diplomate to restrict his activities to his own field. The medical student must decide whether he desires this or prefers to enter immediately into practice and to handle a wider range of cases. At present about 15 per cent of physicians are specialists certified by an American Board. It appears likely that this percentage will double in the next twenty years.

The essence of being a medical specialist consists in the differences in behavior between all medical specialists and all medical nonspecialists. The Specialization Level Scale is based on such groups and measures the differences in interests of specialists from physicians in general. About 30 per cent of physicians obtain an A rating on this scale. (See Technical Note C.)

c. If I Am To Specialize, Which Specialty Should I Select?

The four specialty scales for internist, surgeon, pathologist, and psychiatrist measure the extent to which the physician has the interests of men in each of these specialties as distinct from the interests of physicians in general.

It is expected that scales for other specialties will be developed. In the meantime scores on these scales will give some idea of how one would score on other specialties. For example, a medical student interested in orthopedic surgery ought to score like a surgeon and not like a psychiatrist.

---

[1] This men-in-general group is the same as the P_1 (point of reference group used by Strong for his regular vocational interest scales. Approximately 4500 blanks were used and these were selected to be representative of men in the occupations that college men normally enter (5, p. 711f).

Only medical men should be scored on these specialty scales. The scales are so constructed that they presuppose that the men who are tested have the interests of a physician. No meaning can be attached to scores on these scales when the man does not have the interests of a physician.

d. What About Other Activities Within the Medical Profession?

Some men enjoy and seek administrative work, other men dislike such activities and avoid them. For example, Command and Staff medical officers score like public administrators, whereas typical medical practitioners score much lower on such interest scales. It may be expected that a young M.D. with a low score on the Physician Scale and high scores on the Public Administrator and the Office Worker Scales will like administrative work. He will probably either seek such work or will gradually drift into it as time goes on. Similarly a physician with the interests of an aviator might enjoy the work of a flight surgeon.

Other occupational interest scores should be helpful in planning a medical career which includes activities not involving the actual treatment of patients.

The reader interested only in the general results of this project and how these results can be used by an individual might omit Chapters II and III, reading Chapter IV next. Those interested in the procedures employed in this research should read Chapters II and III next.

CHAPTER II

THE DEVELOPMENT OF THE INTEREST SCALES

A. THE PRIMARY PROBLEM

THE primary problem of this research is the differentiation of the medical specialists, i.e., to try to answer the question, "What specialty?" It is evident that the question, "Should I specialize?" should be answered before considering a particular specialty.

Differentiation has reference to differences, not similarities. The essence of being a physician consists in having those characteristics of physicians not found in other men. Physicians possess interests, objectives, goals not common to other men; they have certain skills which others do not have, etc. The sum total of all these differentiating characteristics is what is meant by being a physician. Differentiation is discussed more fully in Technical Note A.

Differentiation of the medical specialists consists of measuring the differences among the several groups of specialists and between specialists and the average physician. We are not considering differences in abilities or aptitudes but, rather, differences in interests and preferences. It is believed that such differences will predict continuing to like the activities of a particular specialty.

This study is restricted to the four specialties of Internal Medicine, Surgery, Pathology, and Psychiatry. The devices used were the Strong Vocational Interest Blank for Men and the Medical Specialists Preference Blank, the latter blank having been specifically developed for this purpose. Briefly, the procedure was to administer these blanks to groups of specialists and to reference populations of men in general and physicians in general. Several kinds of interest scales were then constructed on the basis of differences in proportions of the various groups responding to the different items on the blanks. These scales were then administered to determine their effectiveness in differentiating the several groups.

B. CRITERION AND REFERENCE GROUPS

In this research a surgeon, for example, is defined as a diplomate of the American Board of Surgery. Although it would not be claimed that all doctors practicing surgery are diplomates of the American Board of Surgery, it would generally be conceded that all such diplomates are surgeons. Similarly, internists, pathologists, and psychiatrists were defined as diplomates of their respective American Boards.

In selecting the names for these four criterion groups from the Directory of Medical Specialists (1), the youngest were preferred. In the case of foreign-born doctors, only those receiving their medical education in this country were included. (See Table 5.)

In order to determine how the interests of specialists differ from those of the average physician, it was necessary to establish a group of physicians as a point of reference. The American Medical Association provided a list of 2500 names randomly selected by machine records

TABLE 5
SIZE AND MAXIMUM AGE OF CRITERION GROUPS OF SPECIALISTS

	Internists	Surgeons	Pathologists	Psychiatrists
Number	1000	1000	600	1000
Year of Birth of the Oldest	1910	1907	1900	1905

TABLE 6

NUMBER OF BLANKS USED IN ITEM COUNTS

	Internists	Surgeons	Pathologists	Psychiatrists	Physicians in General
Vocational Interest Blank	400	400	300	400	500
Medical Specialists Preference Blank	300	300	200	300	400

from their list of all physicians in the United States. This physician-in-general reference group will be referred to as P_{med} (Point of reference—medical).

A large number of occupational interest scales have been developed in the past twenty-five years for the Vocational Interest Blank. Most of these scales were based on comparisons with a group of men in the occupations that college men tend to enter. This was considered the correct reference group since these scales were designed primarily for use by college men. This reference group of men in general was based on over 4500 blanks and will be referred to as P_{men}[1] (Point of reference—men).

C. INTEREST AND PREFERENCE BLANKS

The Strong Vocational Interest Blank for Men consists of 400 items, such as occupations, school subjects, amusements, activities, and peculiarities of people. The person being tested responds by indicating whether he likes, dislikes, or is indifferent to each of the items. This blank has been widely used for many years and some fifty occupational scales have been constructed using the P_{men} reference group.

Since our primary objective is to differentiate among the medical specialists and between each specialty group and the average doctor, it was believed that a blank designed specifically for use by the medical profession might be particularly effective. Various hunches or rationales prompted the development of the different types of items. It was postulated that there would be systematic differences among the four types of medical specialists regarding their preferences for type of professional associate, medical school subjects, casual acquaintances, and descriptions of occupational and recreational activities. Biographical items referring to family and personal background were included.

The Medical Specialists Preference Blank was assembled from items tried out with doctors at three Army hospitals. It consists of 171 items, more than half of which are of the forced-choice type which does not permit an "indifferent" response. The development of this blank is described in Technical Note E.

D. COLLECTION OF BASIC DATA

The Vocational Interest Blank was mailed to the four medical specialist groups in September 1949 and to the physicians-in-general group in May 1950. The Medical Specialists Preference Blank was mailed to these same groups in September 1950. The 3600 doctors in the four medical specialist groups returned 48 per cent of the Vocational Interest Blanks and 42 per cent of the Medical Specialists Preference Blanks.

The basic data required for the construction of interest scales are the percentages of each group marking the respective choices under each item. Item counts were made of responses to the Vocational Interest Blank and the Medical Specialists Preference Blank by each of the four medical specialty groups and by the physicians-in-general group. The number of blanks used in each group is shown in Table 6. Other blanks that were returned were used for cross validation as reported in Technical Note F.

---

[1] See footnote 1, Chapter I.

E. THE INTEREST SCALES FOR MEDICAL SPECIALISTS

The methods used in developing the various scales as well as the technical problems involved are described in detail in Technical Note B. In this section only a brief summary will be given of these procedures.

In constructing scales to measure differences between two groups, each item is assigned a weight on the basis of the difference in the percentages of each group responding in a given way. For example, if a larger proportion of the surgeons than physicians in general like to play golf, then a response of "Like" to the item "golf" on the Vocational Interest Blank would be given a positive weight on the Surgeon Scale.

It is obvious that many different types of comparisons can be made, each of which could result in an interest scale. Scales based on different points of reference could be constructed for each of the specialist groups. Scales could use either the Vocational Interest Blank, or the Medical Specialists Preference Blank, or both blanks combined.

A Physician Scale using the Vocational Interest Blank and based on differences between a group of physicians and the men-in-general reference group has been available for many years. A revised Physician Scale using the P_{med} reference group as the criterion was one of the first scales constructed on this project. An individual's score on this scale indicates whether his interests are more like those of physicians or more like those of men in occupations ordinarily entered by college men. This revised Physician Scale is described in Technical Note D and is based on a more recent and representative sample of physicians than was the original scale.

Using only the Vocational Interest Blank, scales were constructed for each of the four medical specialist groups based, first, on the P_{men} (men in general) reference group and, second, on the P_{med} (physicians in general) reference group. As had been expected, the intercorrelations among the four scales based on P_{men} (average $r = .86$) were much higher than among the four scales based on P_{med} (average $r = .28$). That is, when compared to men in general, internists, surgeons, pathologists, and psychiatrists are much more alike than when they are compared to physicians in general. These relationships indicated that only scales based on the physicians-in-general reference group should be used for differentiating medical specialists.

It was found that scales using *both* the Vocational Interest Blank and the Medical Specialists Preference Blank were superior to scales using either blank alone. Interest scales using both blanks and based on differences between each of the four medical specialist groups and the P_{med} (physicians in general) reference group were constructed. These scales differentiated very well between the specialist groups and the physicians in general (average $r_{bis} = .81$) and fairly well among the four specialist groups (average $r_{bis} = .67$).

The poorest differentiation was on the Internist Scale between internists and pathologists ($r_{bis} = .43$) and between internists and psychiatrists ($r_{bis} = .23$). Are the interests of internists very similar to the interests of psychiatrists and pathologists, or is the relative failure to differentiate caused by faulty scales? In order to answer this question, scales were constructed based on direct comparisons of each of the six pairs of medical specialist groups. The results with these scales indicated that it should be possible to differentiate internists from pathologists and from psychiatrists almost as well as among the other specialist groups.

Several attempts were made to improve the differentiation between internists and psychiatrists and between internists and pathologists. As stated above, the four original medical specialist scales were based on differences between each specialist group and the physicians-in-general reference group. A procedure was developed for selecting those items on the original scales which met certain requirements with reference to discriminating among the four specialist groups. These selected-items scales were substantially better than the original scales in differentiating between internists and psychiatrists ($r_{bis} = .58$ vs. .23) and between internists and pathologists ($r_{bis} = .56$ vs. .43).

Unless otherwise indicated, the Internist, Surgeon, Pathologist, and Psychiatrist Scales referred to in this report are these selected-item scales using both the Vocational Interest Blank and the Medical Specialists Preference Blank.

One other scale has been developed. The interests of the internists, surgeons, pathologists, and psychiatrists, considered as a single group of specialists, are contrasted with the interests of the P_{med} (physicians in general) reference group. The resulting scale is called the Specialization Level Scale. Certain preliminary investigations suggest that scores on this scale may identify individuals who will enjoy taking further specialized training and narrowing the field of their professional work. The development of the Specialization Level Scale is described in Technical Note C.

F. INTERCORRELATIONS OF SCALES AND RELIABILITY

One measure of the goodness of interest scales for purposes of discrimination is based on the intercorrelations of such scales. The lower the intercorrelations, the better the scales. A high positive correlation between two scales means a person obtaining a high score on one scale would tend to get a high score on the other scale. That is, scores of individuals on the two scales would tend to be equal. With a correlation of zero there would be no relationship between the scores of individuals on the two scales. A high negative correlation gives the best differentiation since individuals getting high scores on one scale tend to get low scores on the other.

TABLE 7

INTERCORRELATIONS OF MEDICAL SPECIALISTS SCALES ($N = 100$ PHYSICIANS IN GENERAL)

Scale Correlated	Original Scales	Selected-Item Scales
Internist vs. Surgeon	−.04	−.54
Internist vs. Pathologist	.64	.05
Internist vs. Psychiatrist	.72	.04
Surgeon vs. Pathologist	.11	−.17
Surgeon vs. Psychiatrist	−.20	−.46
Pathologist vs. Psychiatrist	.42	.05
Average	.27	−.17

Table 7 gives the intercorrelations of both the original scales and the selected-items scales. The improvement with the selected-item technique is apparent with a reduction of the average intercorrelation coefficient from .27 to −.17. The Surgeon Scale correlates negatively with each of the other three and the other correlations approximate zero.

Estimates of the reliabilities of the scales were obtained by correlating scores on the odd-numbered items with scores on the even-numbered items and correcting by the Spearman-Brown formula. The resulting reliability estimates for the scales on both blanks combined are: Internist Scale .69, Surgeon Scale .79, Pathologist Scale .89, Psychiatrist Scale .89, and Specialization Level Scale .79. These reliabilities are considered acceptable. The problem of reliability is discussed in more detail in Technical Note F.

CHAPTER III

DIFFERENTIATION OF GROUPS

A. STANDARD SCORES AND LETTER RATINGS

IN ORDER to make comparisons between different groups on the various interest scales, it is necessary to have common units of measurement. Standard scores were established for all scales by setting the mean of the criterion group equal to 50 and the standard deviation equal to 10. These standard scores on the different interest scales are thus directly comparable. Any score given in this report will be a standard score unless otherwise indicated.

Letter ratings from A to C have been used for twenty-five years to express occupational interest scores on the Vocational Interest Blank. An A rating corresponds to standard scores of 45 and above, a B+ rating to scores of 40-44, a B rating to 35-39, a B− to 30-34, and a C rating includes scores of 29 and below.

In explaining the meaning of the letter ratings it may be helpful to consider what proportion of the criterion group would obtain each rating. For example, it would be expected that the scores made by surgeons on the Surgeon Scale would be distributed approximately as follows: A 70%, B+ 15%, B 9%, B− 4%, and C 2%. Most of our criterion groups approximate this distribution quite closely.

Both standard scores and letter ratings will be used to describe the differences between the various groups. The use of interest scales for individual guidance will be discussed in Chapter IV of this report.

B. MEASURES OF OVERLAPPING AND CORRELATION

In comparing the distributions of interest scores made by two groups, a measure of the overlapping has been found useful. Overlapping is defined as "the percentage of scores made by one group which could be matched with scores in the other group." Table 8 shows one way of estimating this overlapping and indicates that medical school seniors and nonmedical college students overlap on the Physicians Interest Scale about 39 per cent. A formula by Tilton (9) gives practically the same result and is easier to use. This formula was used in computing all the estimates of overlap in this report.

TABLE 8

OVERLAP OF MEDICAL AND NONMEDICAL STUDENTS ON THE PHYSICIAN INTEREST SCALE

Rating	Per Cent of Seniors in Medical School	Per Cent of Non-medical College Students	Per Cent* of Overlap of Two Groups
A	73	11	11
B+	14	10	10
B	18	13	13
B−	4	13	4
C	1	53	1
Total	100%	100%	39%

* The smaller of the 2 per cents represents scores common to both groups and, hence, the overlap.

This example indicates that the percentage of overlapping can be considerable even with two groups that are quite sharply differentiated. It may be helpful to consider the overlap of other groups on other scales. Engineers overlap on the Engineer Scale with other occupational groups as follows: chemists 65%, physicians 42%, accountants 36%, and life insurance salesmen 10%. Accountants overlap on the Accountant Scale with

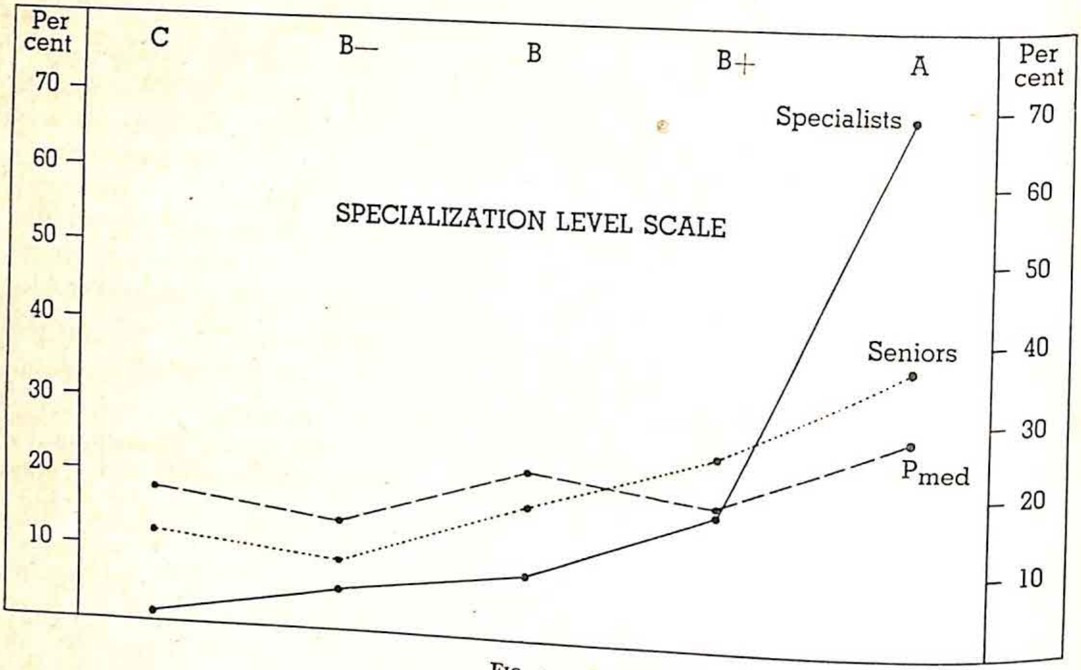


FIG. 1
Differentiation with the Specialization Level Scale
Per cent obtaining each letter rating.
Specialists = All internists, surgeons, pathologists and psychiatrists
Seniors = Medical school seniors
P_{med} = Physicians-in-general reference group

bankers 62%, personnel managers 47%, carpenters 43%, farmers 26%, and ministers 14% (5, p. 109f.).

It should be mentioned that overlap is often expressed as the percentage of a lower-scoring group which equals or exceeds the average of a higher-scoring group. This type of overlap will obviously be much smaller than the one indicated above. It is believed that the procedure used in this study is preferable for describing the total situation.

Some readers may be more familiar with correlation coefficients. The biserial correlation technique measures the relationship between two groups on a single continuous variable such as an interest scale. There appears to be a fairly direct relationship between r_{bis} and overlap in many situations. This relationship can be roughly approximated as follows: r_{bis} .90 = 30% overlap, r_{bis} .70 = 50% overlap, r_{bis} .50 = 70% overlap, and r_{bis} .30 = 90% overlap. Note that the higher the r_{bis} the better is the differentiation between groups.

C. DIFFERENTIATION WITH THE SPECIALIZATION LEVEL SCALE

The Specialization Level Scale reflects the differences between all specialists, considered as a single group, and the physicians-in-general group. This scale was developed to help answer the question, "Should I specialize?"

Figure 1 indicates graphically the percentage of specialists, physicians in general, and medical school seniors obtaining each letter rating on the Specialization Level Scale. The points on the figure representing the percentages in each category are connected by lines to facilitate reading. For example, the ratings of the specialists are distributed as follows: A rating 69%, B+ rating 17%, B rating 8%, B− rating 5%, and C rating 1%.

A larger proportion of the medical school seniors than of the physicians in general have an A or B+ rating on the Specialization Level Scale. This relationship suggests that younger doctors have more of the interests of specialists than do older physicians. (See footnote to Table 27.)

D. DIFFERENTIATION WITH THE FOUR SPECIALIST SCALES

The four scales, Internist, Surgeon, Pathologist, and Psychiatrist, are based on both the Vocational Interest Blank and the Medical Specialists Preference Blank. These scales reflect the differences in interests of groups of internists, surgeons, pathologists, and psychiatrists as compared to physicians in general and with each other. The effectiveness of these scales in differentiating the various groups will now be examined.

Figure 2 indicates graphically the distribution of internists and physicians in general on the Internist Scale. For example, 68 per cent of the internists and only 23 per cent of the physicians in general have an A rating on the Internist Scale. At the other end of the scale, C ratings were obtained by 25 per cent of the physicians in general and only 3 per cent of the internists. Similar information is given for surgeons, pathologists, and psychiatrists with respect to their own scales. Table 9 describes these comparisons in another manner.

The differentiation of the four medical specialty groups from the physicians-in-general group is good with an average overlap of 42 per cent. If we assume that our specialist scales are reasonably good measures of the interests of the specialists, we can make certain estimates. For example, the proportion of the physicians-in-general group having interests similar to those of internists is larger than the proportion having interests similar to any of the other three groups. This apparent relationship seems reasonable. The activities of internists appear to resemble those of the average physician more than do the activities of surgeons, pathologists, and psychiatrists.

How well do the scales differentiate among the four medical specialist groups? Figures 3 and 4 indicate graphically the percentage of each specialty group obtaining each letter rating on each of the four scales. These charts are very similar in shape to those in Figure 2. This

TABLE 9

DIFFERENTIATION OF THE FOUR SPECIALIST GROUPS FROM PHYSICIANS IN GENERAL

Scale	Groups Compared	% Overlap	Biserial Correlation
Internist	Internists and P_{med}*	48	.73
Surgeon	Surgeons and P_{med}	42	.78
Pathologist	Pathologists and P_{med}	37	.83
Psychiatrist	Psychiatrists and P_{med}	41	.80
	Average	42	.79

* P_{med} = Physicians-in-general reference group.

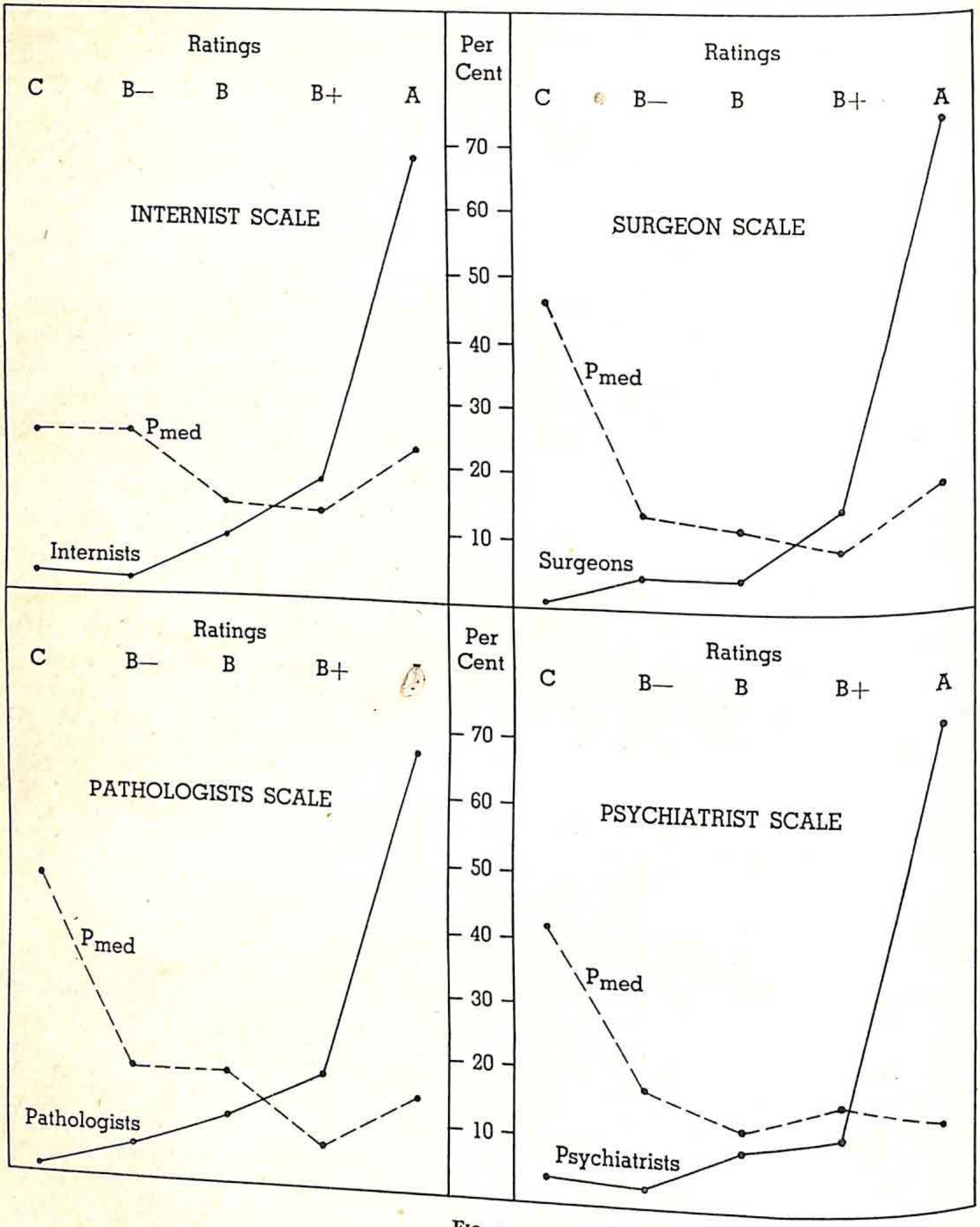


FIG. 2
Differentiation of Specialist Groups from Physicians in General
Per cent obtaining each letter rating

P_{med} = Physicians-in-general reference group

TABLE 10

DIFFERENTIATION AMONG THE FOUR SPECIALIST GROUPS

Scale	Groups Compared	% Overlap*	Biserial Correlation
Internist	Internists vs. Surgeons	31	.90
	Internists vs. Pathologists	61	.56
	Internists vs. Psychiatrists	60	.58
Surgeon	Surgeons vs. Internists	24	.95
	Surgeons vs. Pathologists	31	.89
	Surgeons vs. Psychiatrists	18	.99
Pathologist	Pathologists vs. Internists	53	.66
	Pathologists vs. Surgeons	44	.76
	Pathologists vs. Psychiatrists	42	.79
Psychiatrist	Psychiatrists vs. Internists	57	.63
	Psychiatrists vs. Surgeons	38	.83
	Psychiatrists vs. Pathologists	46	.75
	Average	42	.78

* Overlap computed by Tilton's formula (9).

means that the scales differentiate among the four groups of specialists about as well as they separate each specialist group from the physicians-in-general group. This is substantiated in Table 10 which indicates that the average overlap and average biserial correlation coefficient are about the same as in Table 9.

Surgeons appear to be the most homogeneous and unique of the four medical specialties. The overlap of the interests of internists with those of psychiatrists and pathologists is the largest. However, even here the differentiation is substantial. A biserial correlation of .56 (the lowest in Table 10) represents a high validity for most testing procedures. The average biserial correlation of .78 represents very high effectiveness. (See Technical Note F for consideration of cross-validation data.)

Another way of estimating the effectiveness of these scales is to consider how well they assign specialists to their own specialty. For example, what proportion of the internists will score higher on the Internist Scale than on the Surgeon Scale, what proportion will be tied, and what proportion will show more of the interests of surgeons? Such comparisons are indicated in Figure 5. It should be noted that the "In Error" percentage undoubtedly contains some who actually should have been in another field on the basis of interests.

E. DIFFERENCES AMONG THE FOUR SPECIALIST GROUPS ON OTHER OCCUPATIONAL SCALES

Further differences between the four specialist groups are indicated by the scores on other occupational scales on the Vocational Interest Blank. These scales have been widely used for vocational guidance and were developed on the basis of differences between criterion groups in these occupations and the P_{men} (men in general) reference group. Average scores made by each of the specialist groups are shown in Table 11.

Specialists average slightly higher on the Physician Interest Scale than do

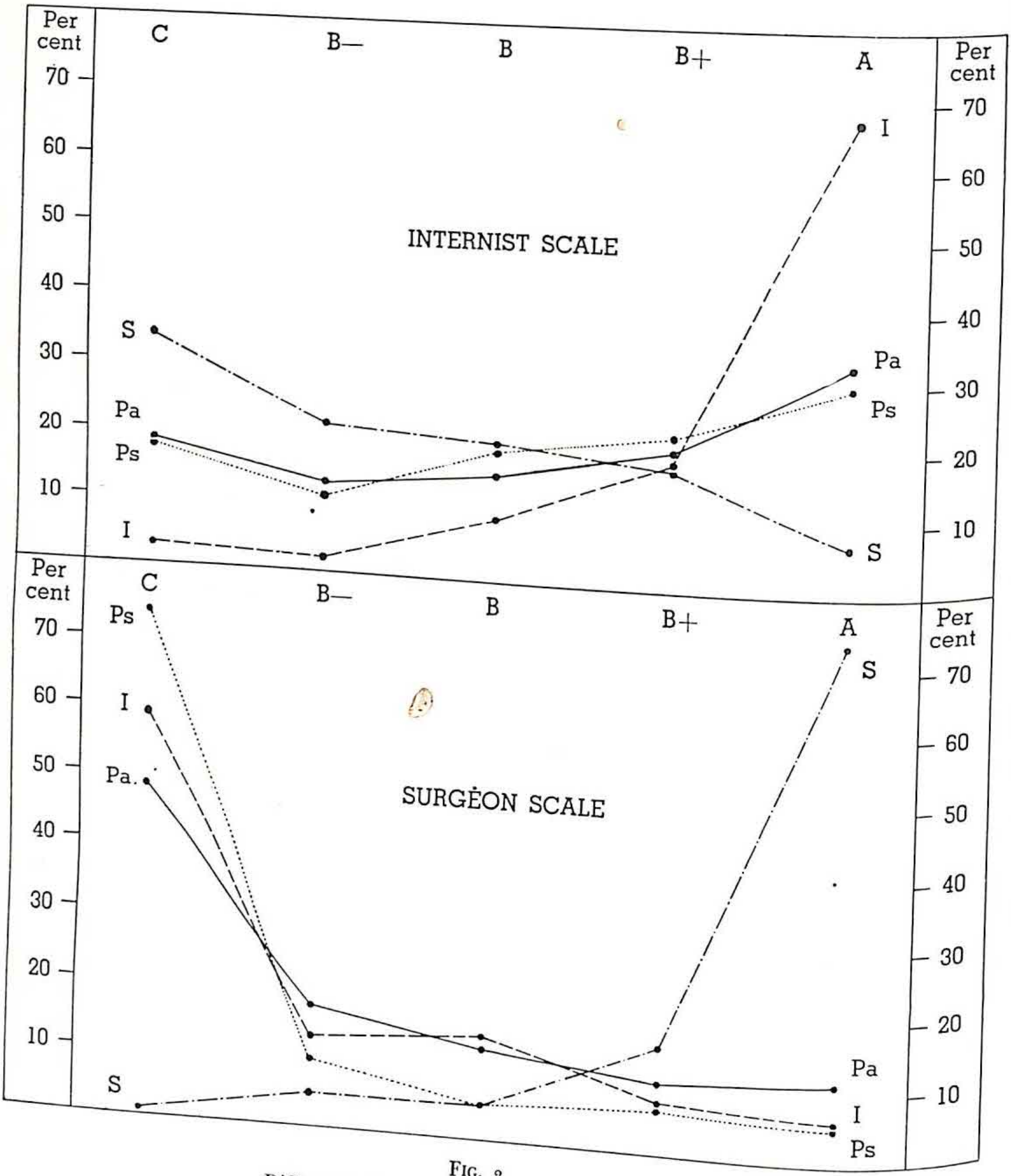


Fig. 3
Differentiation among Specialist Groups with Internist and Surgeon Scales
Per cent obtaining each letter rating

I = Internists, $N = 291$
S = Surgeons, $N = 246$

Pa = Pathologists, $N = 124$
Ps = Psychiatrists, $N = 202$

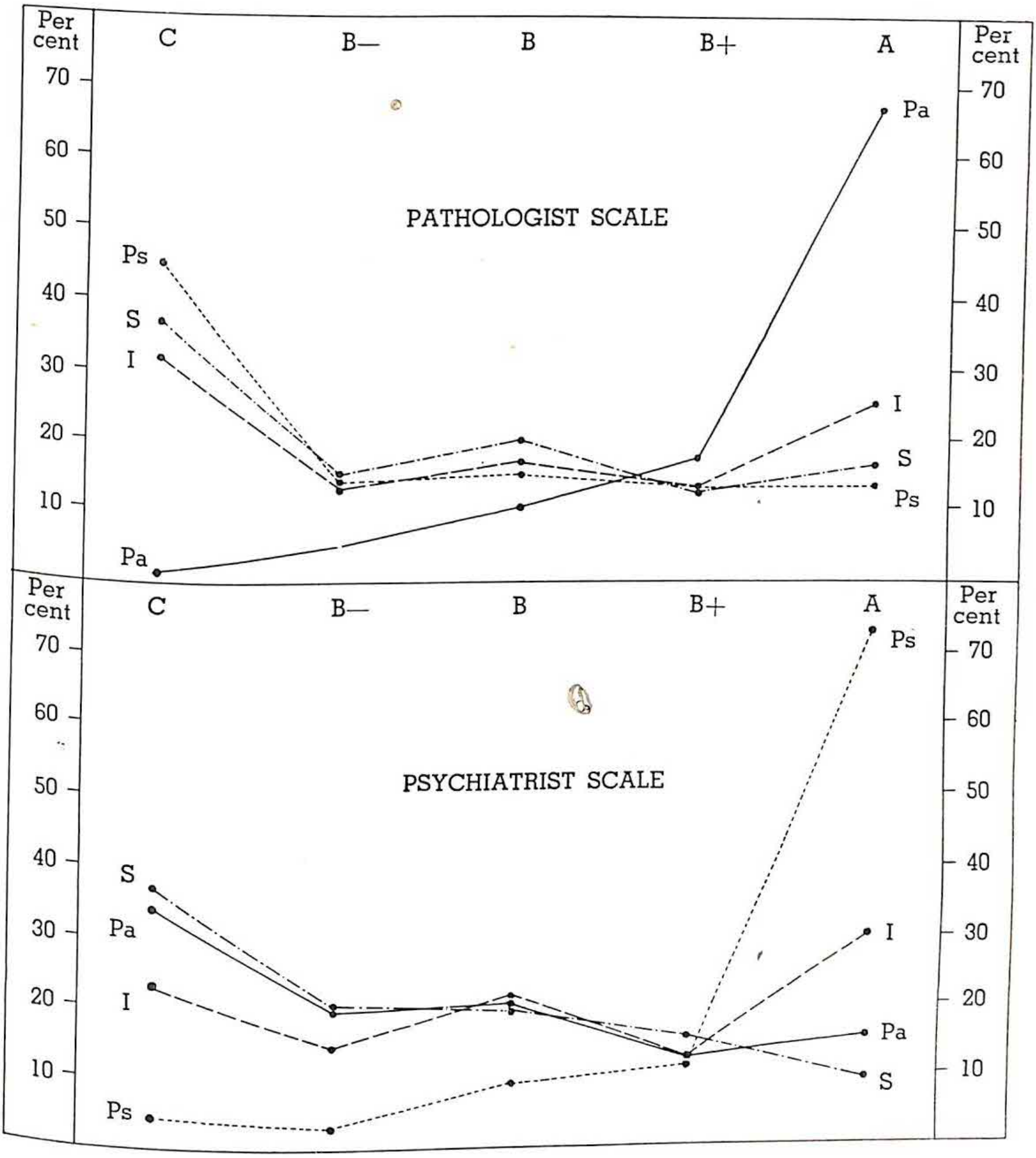


FIG. 4

Differentiation among Specialist Groups with Pathologist and Psychiatrist Scales

Per cent obtaining each letter rating

I = Internists, $N = 291$
S = Surgeons, $N = 246$

Pa = Pathologists, $N = 124$
Ps = Psychiatrists, $N = 202$

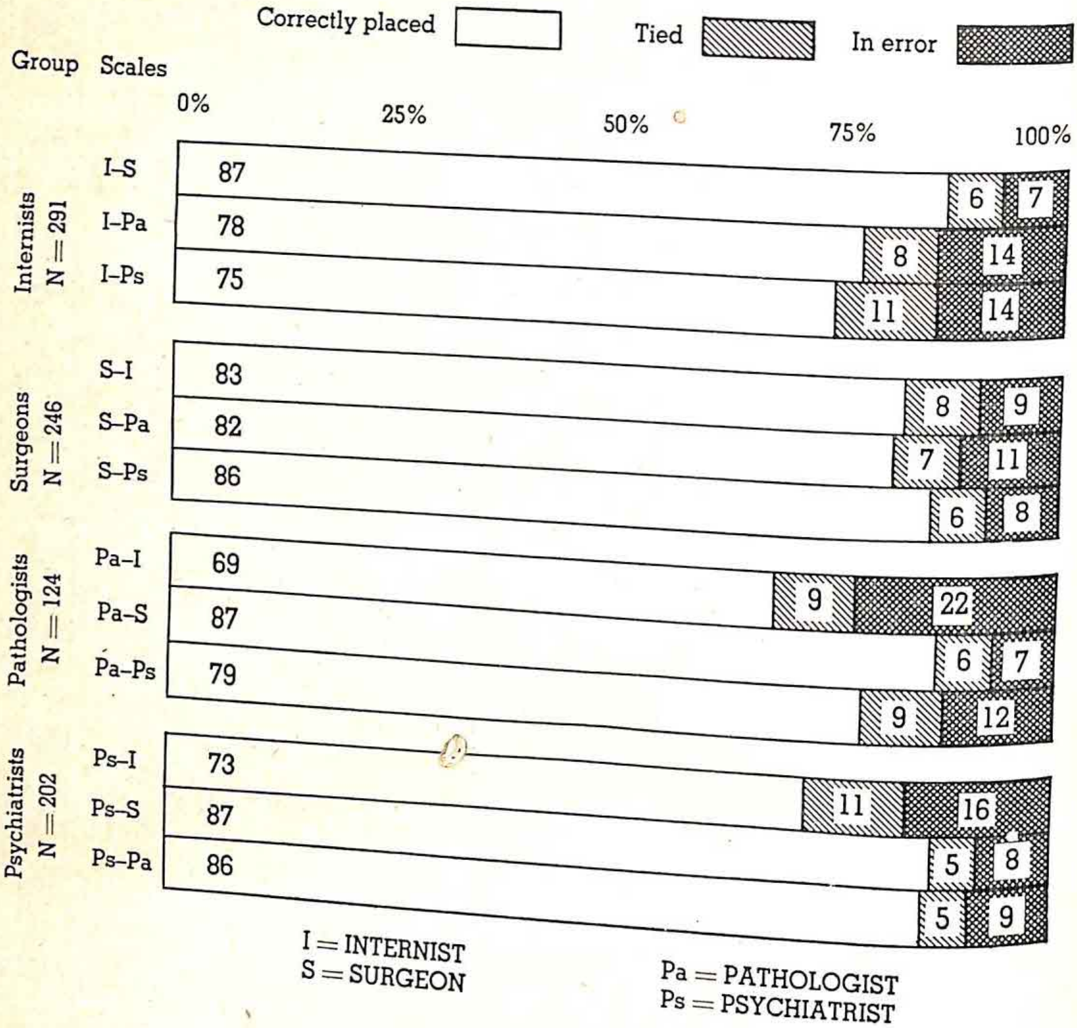


FIG. 5
Per Cent of Each Specialist Group Correctly Placed by Comparison of Scores on Two Scales

physicians in general. They appear to have few of the interests of office workers or life insurance salesmen. Psychiatrists tend to have more of the interests of public administrators, personnel managers, and psychologists than the other three specialists. These relationships seem reasonable, as does the fact that the interests of pathologists are more like those of chemists than are those of the other specialists.

F. RESULTS WITH GROUPS OF OTHER SPECIALISTS

Can the scores of other types of specialists on the medical specialist scales be predicted on the basis of apparent similarity of activities? It was postulated, for example, that orthopedic surgeons, urologists, and neurological surgeons would score higher on the Surgeon Scale than on any of the other three scales. On the other hand, it was believed that

TABLE 11

MEAN SCORES OF SPECIALIST GROUPS ON OTHER OCCUPATIONAL SCALES*
(Based on Vocational Interest Blank)

Scales	Internists	Surgeons	Pathologists	Psychiatrists
Physician (rev.)	52	52	52	51
Public Administrator	42	40	41	46
Office Worker	24	24	22	23
Personnel Manager	32	31	29	39
Math-Science Teacher	37	36	38	37
Psychologist	41	36	42	46
Chemist	41	40	46	39
Life Insurance Salesman	26	27	22	28
Author-Journalist	37	34	37	37
Aviator	31	35	33	31

* Standard Deviations roughly approximate 10.
Letter Rating Equivalents: A = 45 and above, B + = 40–44, B = 35–39, B – = 30–34, C = 29 and below.

pediatricians would tend to score high on both Internist and Psychiatrist Scales.

In order to check these statements, blanks were mailed to diplomates in Orthopedic Surgery, Urology, Neurological Surgery, and Pediatrics. The results are shown in Table 12. On the Surgeon Scale, the orthopedic surgeons score 46, the urologists 41, and the neurological surgeons 45. The average scores of these three groups on the other three medical specialist scales are all lower. Also as predicted, the pediatricians score highest on the Internist and Psychiatrist Scales.

These results definitely support the belief that scores on the four medical specialist scales can be used to estimate the interests of other groups of specialists.

TABLE 12

MEAN SCORES OF GROUPS OF OTHER SPECIALISTS*

Scale	Orthopedic Surgeons $N=61$	Urologists $N=82$	Neurological Surgeons $N=43$	Pediatricians $N=86$
Physician (rev.)	51	49	53	51
Specialization Level	44	42	52	49
Internist	31	33	36	44
Surgeon	46	41	45	22
Pathologist	33	33	41	35
Psychiatrist	33	33	35	41
Public Administrator	41	38	41	42
Office Worker	26	27	23	26
Personnel Manager	31	29	32	32
Math-Science Teacher	40	38	36	40
Psychologist	34	32	38	41
Chemist	39	38	43	39
Life Insurance Salesman	24	26	24	25
Author-Journalist	30	31	34	34
Aviator	39	36	35	31

* Standard Deviations roughly approximate 10.
Letter Rating Equivalents: A = 45 and above, B + = 40–44, B = 35–39, B – = 30–34, C = 29 and below.

TABLE 13

MEAN SCORES OF GROUPS OF NONSPECIALISTS*

Scales	Physicians in General $N=100$	Medical School Seniors $N=150$	Army Interns $N=186$	Army Command and Staff Medical Officers $N=67$
Physician (rev.)	50	51	49	35
Specialization Level	39	42	44	42
Internist	36	34	30	30
Surgeon	31	31	31	32
Pathologist	30	27	28	32
Psychiatrist	33	38	39	38
Public Administrator	36	39	42	48
Office Worker	24	28	31	36
Personnel Manager	27	32	33	41
Math-Science Teacher	34	41	43	37
Psychologist	33	36	34	27
Chemist	37	36	36	28
Life Insurance Salesman	28	28	28	31
Author-Journalist	35	32	29	27
Aviator	32	38	39	33

* Standard Deviations roughly approximate 10.
Letter Rating Equivalents: A = 45 and above, B+ = 40–44, B = 35–39, B − = 30–34, C = 29 and below.

G. RESULTS WITH GROUPS OF NONSPECIALISTS

In addition to the physicians-in-general reference group there are three groups of nonspecialists for which data are available. Approximately 750 seniors in fifteen medical schools completed the blanks in the spring of 1951. These blanks were collected primarily for the purpose of a long-time follow-up study. However, a representative sample of 150 blanks (every fifth blank) has been scored. All interns in Army hospitals also completed the blanks in June 1951. A group of officers in the Army Medical Service who are performing primarily Command and Staff functions was identified by the Office of the Surgeon General; blanks have been returned by 67 of this group.

The mean scores of these nonspecialist groups are given in Table 13. The differences between the Army interns and the medical school seniors are very small. This indicates that the young doctors entering the Army for internships are really representative of all medical school graduates with respect to interests.

The differences between the older physicians-in-general group and the younger medical school seniors are not great and may merely represent differences in emphasis in medical education. For example, there is a tendency for the young doctors to have interests more like those of psychiatrists, math-science teachers, and aviators than their older colleagues.

The most striking feature of Table 13 is the tendency of the Command and Staff group to deviate so markedly from the other three groups. The doctors in this group spend little time in actually caring for patients. Most of them have positions of responsibility in the administration of medical facilities or in making

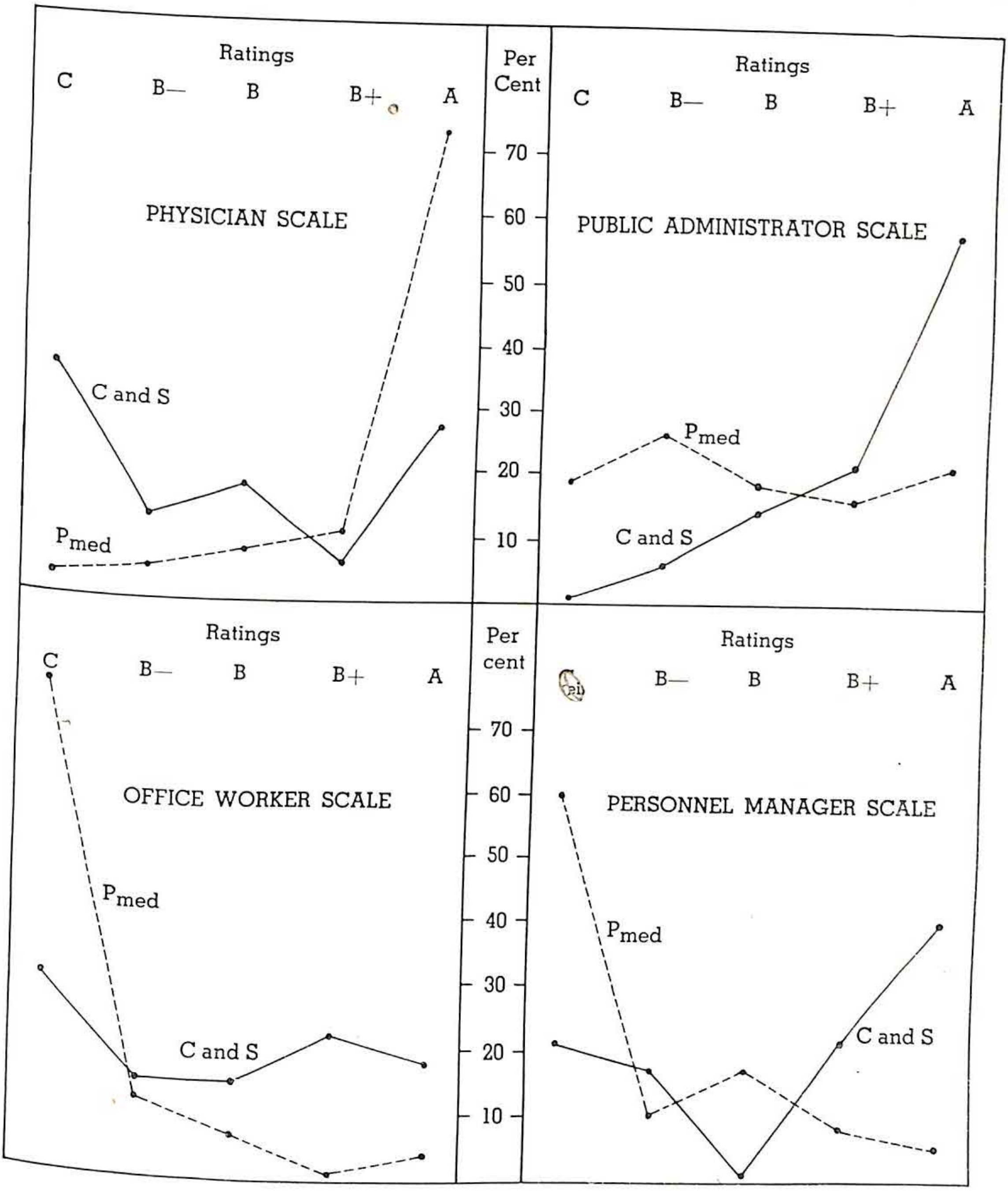


FIG. 6

Comparisons of Command and Staff Group
with Physicians-in-General Group
(Scales on Vocational Interest Blank)
Per cent obtaining each letter rating

C and S = Army Command and Staff Medical Officers Group.
P_{med} = Physicians-in-general reference group.

plans at higher headquarters levels. Some inquiry suggests that most of these doctors actually sought administrative work or at least offered little resistance to such assignments.

Figure 6 presents clearly the large differences between the physicians-in-general group and the Command and Staff group. It appears that only about a third of the Command and Staff group now have the interests of physicians while four out of every five of them get an A or B+ on the Public Administrator Scale.[1] Differences between the two groups on the Office Worker and Personnel Managers scales are equally pronounced and in the direction we would expect.

The differences between groups on the various interest scales which have been discussed in this chapter suggest that interest scores may be used with some confidence by an individual in planning his career. How such results can be reported and used by an individual will be discussed in the next chapter.

---

[1] The data in Table 13 and Fig. 6 concerning the Command and Staff group pertain to their interests as of "now." Whether or not their interest scores have changed from the time they were medical students is, of course, unknown. Table 2 indicates that physician interest scores are fairly stable over a period of 22 years.

Because these men have gone into administrative work, it is possible that their interests have shifted from clinical to administrative activities and the interest scores reflect such a change. Most data do not support this view but rather that existing interests are reflected in occupational activities.

CHAPTER IV

PRACTICAL USE OF INDIVIDUAL SCORES

A. REPORTING MEDICAL INTERESTS

THE man who completes the Vocational Interest Blank and the Medical Specialists Preference Blank wants an answer to one or more of the following four questions:

a. Should I be a physician?
b. If so, should I specialize?
c. If so, which specialty?
d. What about other activities within the medical profession?

The Report of Medical Interests will assist an individual in answering these questions by providing ratings on various scales (copy in Appendix). This report form has undergone several revisions after experimental use with physicians, interns, and medical students. It is a compromise between what may be desired by the research man, on the one hand, and by those relatively unfamiliar with psychological tests on the other hand. Two pages of information briefly answer questions which have been asked by medical men regarding their scores.

Scores are reported in terms of the five letter ratings, A, B+, B, B−, and C, as defined in Chapter III. For the individual, these ratings may be considered as expressions of the probability that he has the interests of men in the listed occupation. An A rating represents a high probability and gives a "Yes" answer to the question, "Does he have such interests?" Similarly a B+ rating means "Probably," a B means "Possibly," a B− means "Probably Not," and a C rating means "No, he does not have such interests."[1]

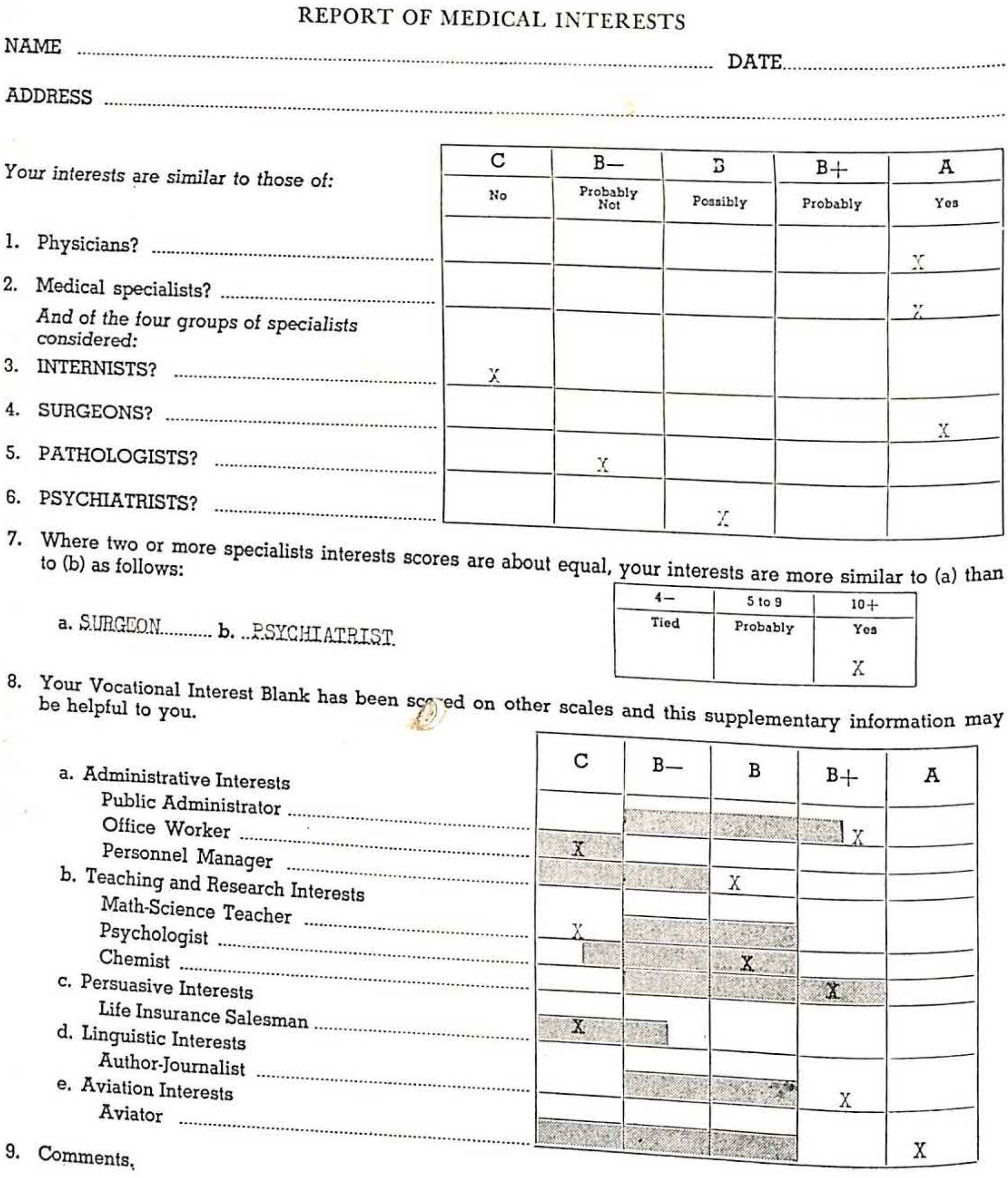
There are always a few men in each occupation who receive a C rating on their own occupational scale. Several analyses of such cases indicate that many of these men do not enjoy their work; others pursue atypical activities, such as superintendent of a hospital or health officer, where they no longer treat patients but do maintain membership in the medical fraternity.

Figure 7 is a copy of the Report of Medical Interests sent to an Army intern. His ratings suggest: (*a*) that he should be a physician, (*b*) that he should specialize, and (*c*) that of the four specialists, he should become a surgeon. His actual scores on SURGEON and PSYCHIATRIST are 52 and 36 respectively. As the difference is greater than 10, an "X" is recorded under "10+, yes" in No. 7, indicating that the SURGEON score is definitely higher than the PSYCHIATRIST score. His scores on the supplementary occupational interest scales (No. 8) approximate those of the average physician; i.e., they fall within or very close to the shaded areas, except on aviator interest, on which he obtains an A rating.

The shaded areas under No. 8 assist in interpreting interests in other activities within the medical profession. For some purposes it is well to consider how one's interests compare, not with men engaged in the ten occupations, but with doctors alone. The shaded areas provide the basis for such comparisons. Twenty-five per cent of physicians score above the shaded area, 50 per cent score within the shaded area, and 25 per cent below this

---

[1] There is substantial evidence that the higher a man's vocational interest score the greater the probability that he has the interests of men in that occupation (5). The use of words such as "Yes," "Probably," and "Possibly" to represent intervals along the gradient of interest scores has been found useful in understanding these scores.



REPORT OF MEDICAL INTERESTS

NAME .... DATE ....

ADDRESS ....

Your interests are similar to those of:	C No	B— Probably Not	B Possibly	B+ Probably	A Yes
1. Physicians?					X
2. Medical specialists?					X
And of the four groups of specialists considered:					
3. INTERNISTS?	X				
4. SURGEONS?					X
5. PATHOLOGISTS?		X			
6. PSYCHIATRISTS?			X		

7. Where two or more specialists interests scores are about equal, your interests are more similar to (a) than to (b) as follows:

a. SURGEON .... b. ..PSYCHIATRIST

4— Tied	5 to 9 Probably	10+ Yes
		X

8. Your Vocational Interest Blank has been scored on other scales and this supplementary information may be helpful to you.

	C	B—	B	B+	A
a. Administrative Interests					
Public Administrator		▨	▨	▨ X	
Office Worker	▨ X				
Personnel Manager	▨	▨	X		
b. Teaching and Research Interests					
Math-Science Teacher	X	▨	▨		
Psychologist	▨	▨	▨ X		
Chemist		▨	▨	▨ X	
c. Persuasive Interests					
Life Insurance Salesman	▨ X	▨			
d. Linguistic Interests					
Author-Journalist		▨	▨	X	
e. Aviation Interests					
Aviator	▨	▨	▨		X

9. Comments.

FIG. 7
Medical Interests of an Army Intern

area. (It should be noted that this use of shaded areas is different from the range of chance scores indicated on the regular Report on Vocational Interest Blank.)

As an example, if a man's rating on office worker is a B, it means that *possibly* he has the interests of office workers; but he has more of the interests of office workers than the great majority of physicians. Such a score suggests that he may like administrative work, or at least dis-

TABLE 14

AVERAGE RATINGS OF OTHER SPECIALISTS ON MEDICAL SPECIALISTS SCALES

Scales	Orthopedic Surgeons $N=61$	Urologists $N=82$	Neurological Surgeons $N=43$	Pediatricians $N=86$
Internist	B−	B−	B	B+
Surgeon	A	B+	A	C
Pathologist	B−	B−	B	B−
Psychiatrist	B−	B−	B	B+

like it less than most doctors. All administrators spend much of their time handling letters and report forms, the work of an office man.

It should be noted that scores on the four medical specialist scales (No. 3, 4, 5, and 6) should not be considered unless one has the interests of a physician. These scores are meaningless if one rates C or B− on the Physician Scale and questionable if one has a B rating. Since these scales were based on differences between groups of medical specialists and physicians in general, they presume a real interest in the medical profession as evidenced by a high score on the Physician Interest Scale. The meaning of scores on these scales for other groups is extremely problematical.

The Specialization Level Scale (No. 2) appears to be most useful in those cases having high ratings on No. 1 and low ratings on No. 3, 4, 5, and 6. Such scores suggest that one has the interests of medical specialists but not of one of the four for which scales have been constructed. Since there are some 32 recognized specialties in medicine, it would appear advisable to investigate carefully some of the other specialties.

How well will scores on the four specialist scales predict interest in other related fields? The specialties of urology and orthopedic surgery, for instance, appear to have much in common with the field of general surgery. Similarly cardiology and dermatology appear closely related to internal medicine. It would appear reasonable to expect the interest scores in the four major specialty fields of internal medicine, surgery, pathology, and psychiatry each to represent the areas of related activities fairly well.

Data for four other types of specialists are given in Table 14. As was expected, the orthopedic surgeons, urologists, and neurological surgeons definitely tend to score higher on the Surgeon Scale than on the other three medical specialist scales. The higher average ratings of the pediatricians on the Internist and Psychiatrist Scales had been predicted on the basis of the professional activities of this group. This evidence, while not conclusive for all types of specialists, certainly supports the belief that a doctor considering any one of the many medical specialties would be aided by considering his results on the scales now available.

B. INTEREST PATTERNS OF GROUPS

Average ratings of physicians in general, senior medical students, Army interns, and Command and Staff officers in the Army Medical Service are indicated in Table 15. The first three groups have average ratings of A on physician interest, higher than on any other scale, which is to be expected. The Command and Staff doctors score only 35 on the Physician Scale, barely a B rating, which is indicative of the fact that

TABLE 15

AVERAGE RATINGS OF NONSPECIALIST GROUPS ON REPORT OF MEDICAL INTERESTS

Interests Similar to—	Physicians in General $N=100$	Medical School Seniors $N=150$	Army Interns $N=186$	Army Command and Staff Medical Officers $N=67$
1. Physicians	A	A	A	B
2. Medical specialists	B	B+	B+	B+
Specialists				
3. Internists	B	B−	B−	B−
4. Surgeons	B−	B−	B−	B−
5. Pathologists	B−	C	C	B−
6. Psychiatrists	B−	B	B	B
8. Other occupational scores				
a. Administrative Interests				
Public Administrator	B	B	B+	A
Office Worker	C	C	B−	B
Personnel Manager	C	B−	B−	B+
b. Teaching and Research				
Math-Science Teacher	B−	B+	B+	B
Psychologist	B−	B	B−	C
Chemist	B	B	B	C
c. Persuasive Interests				
Life Insurance Salesman	C	C	C	B−
d. Linguistic Interests				
Author-Journalist	B−	B−	C	C
e. Aviation Interests				
Aviator	B−	B	B	B−

TABLE 16

PER CENTS HAVING INTERESTS TO SPECIALIZE BASED ON A RATINGS ON PHYSICIAN SCALE, SPECIALIZATION LEVEL SCALE, AND ON ONE OF THE SPECIALIST SCALES

Group	Type of Specialist				Total
	Internist	Surgeon	Pathologist	Psychiatrist	
Medical School Seniors	6	3	5	11	25%
Army Interns	3	2	4	16	25%
Command and Staff Officers	0	3	3	3	9%

they differ from other physicians.[2]

All four groups score about the same on the Specialization Level Scale (No. 2) and on the four specialist scales. The differences between the groups on the other occupational scales are in the directions that one would expect.

How many of each of these groups should become diplomates in each specialty on the basis of their individual interest? This is a crucial question. The answer to this question depends upon what standards are used in determining whether a man should specialize or not. If we require A ratings in physician and specialization interest and an A in one of the four specialist scales, then we have the data in Table 16. The results, if we lower the requirement to include those with either A or B+ ratings on the three scales, are shown in Table 17.

As was to be expected, the Command and Staff group contains a much smaller

---

[2] See footnote 1, Chapter III.

TABLE 17

PER CENT HAVING INTERESTS TO SPECIALIZE BASED ON A OR B+ RATINGS ON PHYSICIAN SCALE, SPECIALIZATION LEVEL SCALE, AND ON ONE OF THE SPECIALIST SCALES

Group	Type of Specialist				Total
	Internist	Surgeon	Pathologist	Psychiatrist	
Medical School Seniors	15	10	6	22	53%
Army Interns	8	11	6	28	53%
Command and Staff Officers	1	4	6	8	19%

proportion of doctors who should specialize than do the other two groups. On the basis of interests alone, somewhere between one-quarter and one-half of the medical school seniors and Army interns should seriously consider specialty training. When other factors in addition to interest are considered, these proportions will be reduced appreciably.

It appears that a substantially larger proportion of the young doctors represented by these two groups have the interests of psychiatrists than of the other three specialists. Is this relationship caused by a faulty psychiatrist scale, or does it represent a real difference in interests of young doctors as compared to older ones? Reference to Table 9 indicates that both the Internist and the Surgeon Scales show less differentiation between the criterion group and the physicians-in-general group than does the Psychiatrist Scale. Since differentiation is one measure of the effectiveness of a scale, the Psychiatrist Scale is considered satisfactory. Table 13 shows the average score on the Psychiatrist Scale for the seniors and interns to be about six points higher than for the physicians in general. It thus appears that the younger doctors do tend to have more of the interests of psychiatrists than do the older doctors. Also, Army interns seem to have more of the interests of psychiatrists than do medical school seniors.

C. EXAMPLES OF INDIVIDUAL RECORDS

It may be helpful in understanding the use of these blanks to consider the records of a few individual cases. In Figure 7 was given the record of a man who meets all the interest requirements as to entry into medicine and specialization in surgery. Our data suggest that at least one-quarter of the young doctors will find that their interests indicate they would enjoy the advanced training and the narrowness of professional activities required of a specialist.

Table 18 (Doctor Y) gives the record of a Command and Staff Medical Corps officer with a very low score in Physician

TABLE 18

RATINGS OF TWO COMMAND AND STAFF MEDICAL CORPS OFFICERS

Scale	Doctor *Y*	Doctor *Z*
Physician (rev.)	C	A
Specialization Level	B	B+
Internist	—	B
Surgeon	—	B+
Pathologist	—	C
Psychiatrist	—	B+
Public Administrator	A	A
Office Worker	B+	C
Personnel Manager	A	B+
Math-Science Teacher	B−	B−
Psychologist	C	B−
Chemist	C	B−
Life Insurance Salesman	B	B
Author-Journalist	C	B−
Aviator	C	B

interest, A ratings in Public Administrator and Personnel Manager, and a B+ rating in Office Worker. He does not treat patients and is an outstanding administrator.

Assuming that this man's interest scores have not changed appreciably since he was in medical school,[3] we may ask, should this man have entered medicine? He has long since ceased to practice. But if he had not become a physician, he would not have risen to a prominent administrative position in the Army Medical Service. On the other hand, if he had entered some field other than medicine, he would probably have risen to an administrative position in that other area. Maybe it would have been a more important position than the one he now holds.

Should the Army take a man with this interest record for internship? The answer is "No" if they want him to practice medicine. If the Army wants doctors who are eventually to become administrators, it would seem preferable to take a man with an interest record similar to Doctor Z (see Table 18). This Command and Staff officer has an A rating on both Physician and Public Administrator scales. While he has made a fine record as an administrator, he devotes much of his time to clinical work.

In selecting Army interns a minimum requirement could be established for the Physician Interest Scale. If an A rating had been required, 76 per cent of the present class of interns would have been selected. Of those selected, 41 per cent would also have A ratings in Public Administrator. If the requirement had been an A or B+ rating on the Physician Scale, then 86 per cent of the present group would have been eligible, with 70 per cent of them having an A or B+ rating in Public Administrator.

---

[3] See footnote 1, Chapter III.

CHAPTER V

SUMMARY AND CONCLUSIONS

A LARGE number of medical school graduates enter the Army each year to serve their internship, and many continue through residency training in one of the medical specialties. These doctors should be assigned to the type of training and work for which they are best fitted and in which they are most interested. To aid in making such assignments, the Surgeon General, U. S. Army, requested the "development of improved methods of guiding Army Medical Corps officers into specialty training. Emphasis will be placed on the development of measures of interest, motivation, and other personal factors, rather than on aptitudes or abilities."

This project was conceived primarily from a guidance, not a selection, point of view. It was recognized that any improved methods developed for guiding Army Medical Corps officers into specialty training would be equally useful to young civilian doctors planning their professional careers.

There are four questions that must be answered by a young man interested in a medical career, namely: (*a*) Should I be a physician? (*b*) If so, should I specialize? (*c*) If so, what specialty? (*d*) What about other activities within the medical profession? One important aspect of these questions involves job satisfaction and continuing high motivation in the work. Obviously, there are other aspects of these questions, such as aptitudes, abilities, and financial support, which were not considered in this research.

The results of several extensive follow-up studies of Stanford students extending up to 22 years were presented as evidence that scores on the Strong Vocational Interest Blank do have considerable permanence and predict vocational behavior quite well. The Physician Interest Ratings of 670 Stanford students, of whom 108 were physicians twenty years later, were considered. Those students with an A rating in physician interest had 53 chances in 100 of becoming a physician; with a B+ rating, 20 chances; B rating, 12 chances; B− rating, 11 chances; and with a C rating, only 2 chances in 100. These studies lead one to expect that similar measures of interests of medical specialists would be useful in planning a medical career.

The primary problem of this research was to develop measures of interest and other types of preference which would differentiate the various types of medical specialists. Based on the evidence indicated above, it was believed that such measures would be predictive of continuing to enjoy working in a particular specialty. Only the four major specialties of Internal Medicine, Surgery, Pathology, and Psychiatry were studied.

The criterion groups consisted of diplomates certified by the American Boards in these four specialties; i.e., a surgeon was defined as a diplomate of the American Board of Surgery. A physicians-in-general reference group was randomly selected by the American Medical Association from their lists of all physicians in this country.

Two blanks were used for measuring interests. The Strong Vocational Interest Blank has been used for 25 years to measure the vocational interests of many different occupational groups. The Medical Specialists Preference Blank was developed specifically for use by the medical profession. Many of the items on this blank relate directly to the practice of medicine and more than half are of the forced-choice type, which does not permit an "indifferent" response. These items were first tried out on doctors in three large Army hospitals.

Each of these two blanks was mailed at different times to 1000 diplomates in each of the specialties of Internal Medicine, Surgery, and Psychiatry, and to 600 diplomates in Pathology. The initial mailing to the physicians-in-general group was 2500, but this was greatly reduced by the subsequent elimination of women and doctors over 55 years of age in the original listing. The samples on which item counts were made approximated 400 for each of the criterion groups. Data were also collected on smaller samples of the following: orthopedic surgeons, urologists, neurological surgeons, pediatricians, medical school seniors, Army interns, and Army Command and Staff medical officers.

Several types of scales were constructed, based on differences in the proportions of the various groups making different responses to each item. Weights varying from −4 to +4 were assigned to each item. There were scales which contrasted each specialist group (*a*) with men in general, (*b*) with physicians in general, or (*c*) directly with each other specialist group. Scales were constructed using each of the blanks separately or based on both blanks combined.

A selected-item scale using both blanks, which maximized both the differentiation of the criterion group from the physicians-in-general group and from the other criterion groups, was found to be most effective. The results presented below are based on four such scales, one for each of the following types of specialists: Internist, Surgeon, Pathologist, and Psychiatrist.

Two other scales were developed. The item counts of the physicians-in-general group on the Vocational Interest Blank were used to construct a revised Physician Scale for use with that blank. This scale contrasts the interests of physicians with men in general and assists in answering the question, "Should I be a physician?" (See Technical Note D.)

A Specialization Level Scale was constructed based on the differences between all of the specialists, considered as a group, and the physicians-in-general group. This scale was designed to help answer the question, "Should I specialize?" (See Technical Note C.)

The results, in terms of differentiation between groups, can be summarized as follows:

1. A larger proportion of the medical-school seniors than of the physicians-in-general group have an A or B+ rating on the Specialization Level Scale. This relationship suggests that younger doctors have more of the interests of specialists than do the older physicians.

2. The Internist, Surgeon, Pathologist, and Psychiatrist Scales each differentiate the criterion group from physicians-in-general group exceptionally well (average $r_{bis} = .79$, range .73 to .83). These four scales also differentiate among the four specialist groups very well (average $r_{bis} = .78$, range .56 to .99). Surgeons appear to be the most homogeneous and unique of the four medical specialists. The overlap of the interests of internists with those of psychiatrists and pathologists is the largest.

3. It was postulated that orthopedic surgeons, urologists, and neurological surgeons would average higher scores on the Surgeon Scale than on the other

three scales. It was also believed that pediatricians would score highest on the Internist and Psychiatrist Scales. These hypotheses were definitely supported by the data on samples of these four specialists. In every case the average ratings on the other scales were at least two ratings lower than on the scales postulated to be high.

4. The three groups of nonspecialists —that is, medical school seniors, Army interns, and Army Command and Staff medical officers—average a B rating or less on the four medical specialist scales. These groups, as a whole, would not be expected to show the interests of a specialist group.

5. The Vocational Interest Blanks were also scored on several of the regular interest scales which contrast the interests of occupational groups with men in general. The psychiatrists tend to have more of the interests of public administrators, personnel managers, and psychologists than the other three specialists. The interests of pathologists are more like those of chemists than are the other specialties.

6. Among the groups of nonspecialists, the Army Command and Staff medical officers group deviates markedly on the regular interest scales from the other groups. It appears that only about a third of the Command and Staff group now have the interests of physicians, while four out of five get an A or B+ on the Public Administrator Scale. Average scores of this group on the Office Worker and Personnel Manager Scale are substantially higher than for the other medical groups. The doctors in this group spend little time in actually caring for patients. Most of them have positions of responsibility in the administration of medical facilities or in making plans at higher headquarters.

The Report of Medical Interests was designed for use in reporting an individual's scores in such a way as to be most useful in his own planning. Scores are reported in terms of the five letter ratings, A, B+, B, B−, and C, which may be considered as expressing the probability that he has the interests of men in the listed specialty or occupation. An A rating means "Yes," a B+ rating means "Probably," a B means "Possibly," a B− means "Probably Not," and a C rating means "No, he does not have such interests."[1] The use of these results by an individual will be considered with reference to the four questions he must answer.[2]

a. Should he be a physician? An A or B+ rating on the Physician Scale supports his belief that he would enjoy the practice of medicine. A rating of B or lower suggests he should consider other occupations.

b. If so, should he specialize? A high score on the Specialization Level Scale indicates he has the interests of a specialist; i.e., he would probably enjoy the narrowing of his area of professional activities as well as the further training required of a specialist.

c. If so, what specialty? An A or B+ on the Internist, Surgeon, Pathologist, or Psychiatrist Scale indicates he probably has the interests of that type of specialist, *provided he has at least an A or B+ on the Physician Scale.* Since these scales measure the differences between specialists and physicians in general, the scores have no meaning unless one has the interests of a physician. While scales are available for only these four medical specialists, it is believed that these results can be used with some confidence to estimate how one's interests would probably correspond with those of other types of specialists. For example, cardiology and dermatology appear to be closely related to internal medicine. Such use of these ratings is definitely supported by the findings for orthopedic surgeons, urologists, neurological surgeons, and pediatricians.

---

[1] See footnote 1, Chapter IV.

[2] Other factors besides interest, such as aptitude, ability, and social need, must also be considered.

d. What about other activities within the medical profession? Some physicians devote little or no time to the treatment of patients but, rather, engage in such activities as teaching, research, hospital administration, preventive medicine, marketing medical supplies, etc. The scores on the ten occupational scales on the Vocational Interest Blank may be helpful in considering many of th[illegible]rious activities within the medical profession. In addition to letter ratings indicating how one's interests compare with those of men in these occupations, shaded areas are marked to indicate the range of scores made by the middle 50 per cent of the physicians-in-general group. For example, if his rating on the Office Worker Scale is a B, he *possibly* has the interests of office workers; but he has more of the interests of office workers than the great majority of physicians.

Certain problems of methodology are discussed in detail in the six Technical Notes. The principles of differentiation and point of reference are elaborated in Technical Note A. A new method of selecting interest items so as to maximize differentiation among several criterion groups as well as between the criterion group and the reference group is presented in Technical Note B. The other four Technical Notes discuss other aspects of the research.

Technical Note A

DIFFERENTIATION AND POINT OF REFERENCE

1. *Differentiation*

Differentiation has reference to differences, not similarities. The essence of being a physician consists of having those characteristics of physicians not found in other men. The sum total of all of these differentiating characteristics is what is meant by being a physician.

Differences exist between two or more things. In this research the things are groups of men. Such groups may be likened to spheres. The center of each sphere is the mean score of a test; the radius is expressed in terms of the standard deviation of the scores. Such spheres are not to be thought of as isolated in space but, rather, as overlapping one another. The better the differentiation the farther apart are the centers of the spheres and the less the spheres overlap.

On several occasions lists have been prepared of the characteristics of successful men in various occupations such as presidents of corporations, foremen, salesmen, etc. Such lists contain almost identical items. Little is gained by use of such lists, for all men, according to these lists, are far more similar than different. If an understanding of successful presidents is desired, the characteristics that distinguish such men from unsuccessful presidents, or vice-presidents, or from foremen, must be obtained.

In differentiating two groups it is necessary, first, to note the characteristics that distinguish them from each other and, second, to use devices to measure these characteristics. If the difference is a matter of stature, a tape is employed; if it is mathematics, an achievement test is used; if it is interests, an interest or preference blank is utilized.

The characteristics that differentiate two groups are present in individuals in both groups but in varying degree. The interest test of masculinity-femininity takes into account both mechanical interests possessed by men more than women and linguistic interests possessed by women more than men. Three lists of differentiating characteristics would need to be employed in the three cases where surgeons are contrasted with all M.D.'s, all college graduates, and all men in the country.

2. *Point of Reference*

If one is asked to comment on some coins, ranging from cents to silver dollars, he tells you that they differ in size, color, value, and metal composition. If he is asked to comment on the contents of a desk drawer, he mentions coins, rubber bands, paper clips, pencils, etc. If the request is made about one's study, the reply is desk, books, chairs, etc. In each case the individual unconsciously reacts in terms of the center of all the things to be noted. In the first case, the cent is smaller than the dollar. In the second case, all the coins are grouped in contrast to pencils, etc. In the third case, the coins are not mentioned, nor are the other contents of the drawer—they are all grouped into the term "desk," for everything in the room is noted in reference to the room as a whole.

Differentiation of groups of men is concerned not only with the peculiarities of the two groups, but also with points of reference. In differentiating medical specialists they are viewed as deviants from all medical men. They can also be thought of as deviants from all profes-

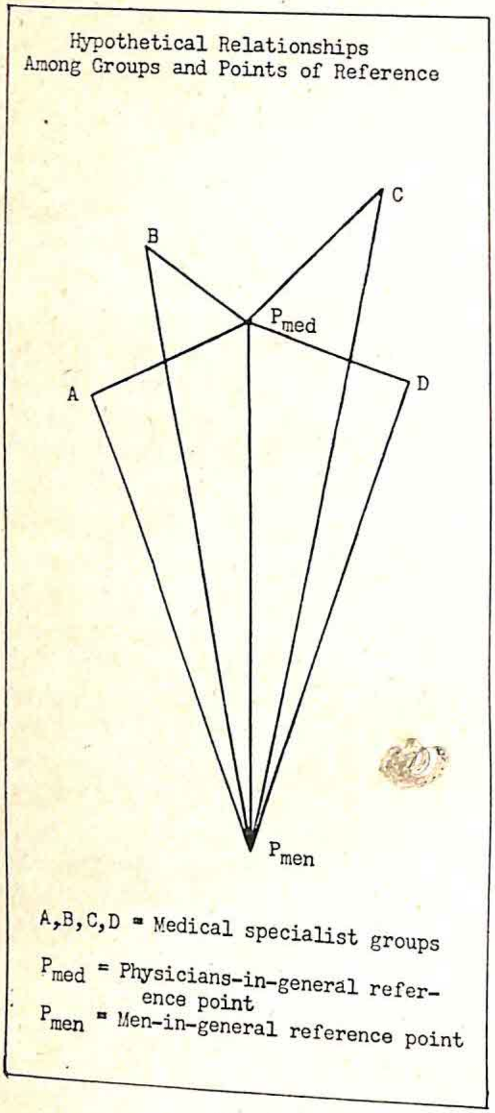

Hypothetical Relationships
Among Groups and Points of Reference
C
B
P_{med}
A
D
P_{men}
A,B,C,D = Medical specialist groups
P_{med} = Physicians-in-general reference point
P_{men} = Men-in-general reference point

FIG. 8

TABLE 19

Correlations Represented by Figure 8

Part 1

Physicians-in-general reference point

Relationship	Angle in degrees	Correlation (angle cosine)
A - P_{med} - B	59	.51
A - P_{med} - C	162	-.95
A - P_{med} - D	133	-.69
B - P_{med} - C	101	-.20
B - P_{med} - D	163	-.96
C - P_{med} - D	64	.43
	Average	-.31

Part 2

Men-in-general reference point

Relationship	Angle in degrees	Correlation (angle cosine)
A - P_{men} - B	10	.99
A - P_{men} - C	32	.85
A - P_{men} - D	40	.77
B - P_{men} - C	22	.93
B - P_{men} - D	30	.87
C - P_{men} - D	7	.99
	Average	.90

sional men, or from all college graduates.

In this research two points of reference are used. The question, "Should this man be a physician?", means "Does he have the interests of a physician to a greater degree than he has the interests of an engineer, lawyer, accountant, and all the other occupations that college men are engaged in?" Since only college men become physicians, we do not concern ourselves with occupations not attractive to college men. The point of reference in this case is an average of the men engaged in the occupations attractive to college men. This point of reference is designated as P_{men}.[1] It may be considered as the center of all the occupations young college men might consider entering. It is also the center of all the young men who might consider medicine as a career.

The second point of reference is used

---

[1] See footnote 1, Chapter I.

in connection with the question, "Which specialty should I enter?" This question can be stated: "Are my interests those of specialists; and if so, in which specialty would I most enjoy the work?" The only persons concerned with this question are practicing physicians, interns, and medical students. The point of reference is accordingly a physician-in-general group, referred to here as P_{med}.

In order to show how the statistics here employed measure the relationships expressed above with regard to P_{men} and P_{med}, Figure 8 is presented. This diagram is not drawn to any scale, nor based on specific data, and is presented solely to give a general idea of certain relationships. P_{men} is shown at the bottom of the figure and P_{med} in the center. The four letters, A, B, C, and D, represent four groups of medical specialists.

When interest scales are constructed differentiating the interests of medical specialists from P_{med}, the scales may be represented by the lines A-P_{med}, B-P_{med}, C-P_{med}, and D-P_{med}. Since each scale represents the *differences* in interests of physicians and a group of specialists, the interests pertinent to medicine are eliminated and the scale represents only the characteristics in which the specialty group differs from the average physician.

Correlations between the interests of two specialist groups—i.e., between A and B, A and C, A and D, B and C, etc.—are, strictly speaking, not between the two types of specialists but involve P_{med} also. The correlations are actually between (*a*) the differences in interests of A and P_{med} and (*b*) the differences in interests of B and P_{med}. In such a situation the correlation coefficient is equal to the cosine of the central angle, i.e., to the angle A-P_{med}-B. The angles between the six combinations of A, B, C, and D and the corresponding cosines of these angles, which equal the correlations, are given in Part 1 of Table 19.

If interest scales are constructed differentiating the interests of medical specialists not from P_{med} but from P_{men}, the scales may be represented by the lines A-P_{men}, B-P_{men}, C-P_{men}, D-P_{men}. Here the interests peculiar to all men in occupations entered by college men are eliminated. But the interests of physicians as such are not eliminated and all four scales represent the differences between P_{men} and the interests of the average physician, modified somewhat by the variations in interests of the four specialties. When Groups A to D are contrasted in terms of P_{men}, the angles are all acute and so the cosines and correlations approximate 1.00, as shown in Part 2 of Ta[illegible]9.

The Physician Interest Scale used to measure the interests of young college men who are considering which occupation to enter is represented in the diagram by the line P_{med}-P_{men}. The angles between P_{med}-P_{men}-A, P_{med}-P_{men}-B, etc. are all very acute angles. Hence the correlations will be high, which is appropriate, since the interests of physicians and of medical specialists are quite similar. The intercorrelations between scores of physicians and scores of the four specialist groups average .96 in the diagram, higher than the intercorrelations among the four specialist groups, when considered in terms of P_{men} (average .90). This relationship appears reasonable since specialists will agree more closely with all physicians than will each specialist agree with other specialists.

Table 19 gives two sets of correlations between specialists (A and D) which are

TABLE 20

INTERCORRELATIONS OF SCALES ON THE VOCATIONAL INTEREST BLANK BASED ON TWO DIFFERENT POINTS OF REFERENCE

($N = 285$ Stanford Seniors)

Scales	Scales (P_{men} Reference Point)				Scales (P_{med} Reference Point)			
	I	S	Pa	Ps	I	S	Pa	Ps
Internist	—	.909	.927	.846	—	.324	.365	.508
Surgeon	.909	—	.899	.792	.324	—	.118	.234
Pathologist	.927	.899	—	.770	.365	.118	—	.110
Psychiatrist	.846	.792	.770	—	.508	.234	.110	—
	Average = .857				Average = .276			

very different. The members constituting the specialist groups are identical and the relationships among them cannot, therefore, have changed. What has changed is the point of reference. In Part 1, the correlations are small or negative; that is, each specialist group appears to differ greatly from the others. This situation is comparable to looking at coins on a table. When coins alone are se[illegible]he rather minute differences among them are noted. In Part 2 the correlations are large and positive; that is, the specialists appear as very similar. This is comparable to seeing coins mixed in with all the items in a desk drawer—they are just coins. Thus, when we consider men in the occupations entered by college men, the internists, surgeons, pathologists, and psychiatrists are just physicians.

As stated above, the diagram in Figure 8 was drawn more or less at random. The specific data given in Table 19 apply to this particular diagram. But the general relationships will hold regardless of the figure used so long as P_{med} is located approximately in the center of A, B, C, and D; and P_{men} is located at some distance from P_{med}.

In contrast to Table 19, Table 20 gives the correlations between the scales, based on actual scores on the Vocational Interest Blank. When the four specialist groups are contrasted with P_{men}, the four scales correlate, on the average, .857 with each other. When the four specialist groups are contrasted with P_{med}, the four scales correlate, on the average, .276 with each other. The lower correlations and hence better differentiation of the scales based on P_{med} as compared to the scales based on P_{men} is readily apparent.

Maximum differentiation is obtained by contrasting groups with a point of reference which is located in the center of the groups. As the point of reference is located farther and farther away from the groups, the differentiation becomes less and less.

Technical Note B

DEVELOPMENT OF THE MEDICAL SPECIALIST SCALES

1. *Introduction*

This Technical Note explains the procedures used in developing the interest scales recommended in this report. In order to understand these procedures it will be necessary for the reader to keep clearly in mind the following four types of scales:

Type of scale	Based on differences between
A	Criterion group and P_{men} (men in general) reference group
BO	Criterion group and P_{med} (physicians in general) reference group
BS	Criterion group and P_{med}, modified by differences *among* specialist groups.
C	Each pair of specialist groups measured directly.

Using scales based on the Vocational Interest Blank only, it is demonstrated in Technical Note A that scales based on P_{med} have much lower intercorrelations than scales based on P_{men}. The type A scales for the medical specialists will, therefore, not be considered further in this section.

2. *Evaluation of the BO Scales*

The primary function of the BO scales is to differentiate each specialist group from P_{med}. This was accomplished very well when both blanks were utilized. Table 21 indicates that the average overlapping is only 39 per cent between specialist groups and P_{med} with a corresponding biserial r of .81.

It is also essential to the effectiveness of these scales that they differentiate among the four groups of specialists. If psychiatrists, for example, have nearly as high scores on the Internist Scale as the internists do, guidance based on such a scale is likely to be quite faulty. Actually, differentiation between pairs of specialists is excellent in 7 of the 12 comparisons with an average overlap of 40 per cent (range 27 to 49). On 3 other comparisons the overlapping ranges from 59 to 64 per cent, which is still fairly good differentiation. But in 2 comparisons the overlapping is 76 and 85 per cent respectively, which represents too small differentiation to be of practical value.

The poorest differentiation involves internists in contrast to psychiatrists and pathologists. It is poorest between these groups on the Internist Scale.

In order to obtain some estimate of what differentiation was possible between pairs of specialists, the C scales were constructed. For this purpose each specialist group was directly contrasted with each of the others. The percentage of overlapping between the groups of specialists on these C scales is indicated in Table 22.[1] These data warrant the expectation that better differentiation can be obtained than resulted from the BO scales and, furthermore, that internists can be differentiated nearly as well as the other three specialists.

3. *Weighting of Part Scores on BO Scales*

Attempts to improve the BO scales proceeded along two different lines. The first attempt involved weighting parts of the blanks so as to give maximum differentiation. The second attempt involved selecting items to accomplish this purpose.

The problem of differentiating quali-

---

[1] It is recognized that the amount of overlapping may be spuriously low. In this situation the actual amount is not as important as the fact that the internists are differentiated relatively as well as the other groups.

TABLE 21

DIFFERENTIATION OF BO AND BS SCALES

($N=100$ for each group)

Scale	Groups Compared	BO Scales				BS Scales			
		Per Cent of Overlap			r_{bis} Both Blanks	Per Cent of Overlap			r_{bis} Both Blanks
		VIB Only	MSPB Only	Both Blanks		VIB Only	MSPB Only	Both Blanks	
Internist	P_{med} vs. I	52	52	49	.71	57	54	48	.73
Surgeon	P_{med} vs. S	46	39	34	.86	52	49	42	.78
Pathologist	P_{med} vs. Pa	44	46	37	.83	43	50	37	.83
Psychiatrist	P_{med} vs. Ps	40	47	38	.83	44	49	41	.80
Average		45	46	39	.81	49	50	42	.78
Internist	I vs. S.	62	37	45	.75	53	30	31	.90
	I vs. Pa	85	68	76	.43	74	61	61	.56
	I vs. Ps	99	72	85	.23	75	55	60	.58
Surgeon	S vs. I	55	30	33	.88	40	31	24	.95
	S vs. Pa	67	45	47	.75	51	40	31	.89
	S vs. Ps	61	21	27	.93	38	25	18	.99
Pathologist	Pa vs. I	71	68	64	.53	59	61	53	.66
	Pa vs. S	53	48	44	.77	48	54	44	.76
	Pa vs. Ps	74	55	60	.59	53	50	42	.79
Psychiatrist	Ps vs. I	64	[illegible]	59	.59	60	62	57	.63
	Ps vs. S	49	[illegible]	35	.86	50	36	38	.83
	Ps vs. Pa	56	31	49	.71	52	47	46	.75
Average		66	49	52	.67	54	46	42	.77

VIB = Vocational Interest Blank
MSPB = Medical Specialists Preference Blank
r_{bis} = Biserial correlation coefficient
P_{med} = Physicians-in-general reference group
I = Internists
S = Surgeons
Pa = Pathologists
Ps = Psychiatrists

tative classes of individuals with a minimum of overlapping when one has several measurements on each individual can be solved statistically with Fisher's discriminant function. The data required for the computation of this statistic consist of intercorrelations of the variables being measured and estimates of the effectiveness of each variable in differentiating between each two classes. The

TABLE 22

DIFFERENTIATION OF C TYPE SCALES

($N=100$ for each group)

Scale	Groups Compared	Percentage of Overlapping
Internist-Surgeon	Internist vs. Surgeon	23
Internist-Pathologist	Internist vs. Pathologist	39
Internist-Psychiatrist	Internist vs. Psychiatrist	37
Surgeon-Pathologist	Surgeon vs. Pathologist	21
Surgeon-Psychiatrist	Surgeon vs. Psychiatrist	24
Pathologist-Psychiatrist	Pathologist vs. Psychiatrist	28
	Average	32%

computational procedures involved in determining the weights to be applied to the variables for maximum differentiation between two classes are relatively simple. For more than two classes the computations become very complex.

Intercorrelations between the 14 parts of the two blanks on each of the four scales were computed. Means and standard deviations for each of the four groups of specialists were computed on each of the 14 parts on each of the four scales. Punched card machine methods were used in making these computations.

A number of solutions using the discriminant function method were worked out between successive pairs of specialties. The effectiveness of using the resulting part-score weights was evaluated by estimating the percentage of overlapping between the various classes.

Some improvement in the discrimination between certain groups of specialists could be accomplished by the use of weighted part scores. However, such improvement required a great reduction in the number of part scores being weighted and poorer differentiation among other groups.

It became apparent that these part scores were not measuring separate variables in the sense of being fairly homogeneous and somewhat independent. The use of differential weights for the several part scores of the two blanks did not appear to be the best method of improving the effectiveness of the scales.

4. *Selection of Items for BS Scales*

It became obvious that occupational items, for example, do not measure a different interest from amusement items. Actually, one combination of occupations and amusements is liked by internists, another combination by surgeons, and so on. The second attempt to improve the scales *viewed each item as a separate test,* just as parts of a battery are viewed as separate tests. As it was impossible to handle 571 items as one does ten or fifteen tests in a battery, some other approach than Fisher's discriminant function had to be devised.

The procedure followed in selecting items for the revised BO scales, called BS scales, is illustrated in Table 23, where four items are considered in detail. For each of the four items the percentages of like, indifferent, and dislike responses to the item by P_{med} and each of the four specialists groups are given. The weights assigned the item are obtained directly from a table which gives every possible combination of 0 to 100 per cent. Thus, in the case of item No. 4, the percentage of 16 for liking by internists is contrasted with the percentage of 21 by P_{med} and the table gives a weight of 0. Similarly the percentage of 23 for indifference by internists is contrasted with the percentage of 30 by P_{med} and the table gives a weight of 1. Since fewer internists are indifferent to the item than P_{med}, the negative sign is attached, so that the weight reads -1. In the same manner a $+1$ weight is assigned dislike since 61 per cent of internists dislike the item in contrast to 49 per cent of P_{med}. Only zero weights are obtained from the table when the responses of surgeons, pathologists, and psychiatrists are contrasted with the responses of P_{med}.

This procedure was used in weighting items for the BO scales. In this case item No. 4 is used on the Internist Scale, but not on the other three BO scales. The weights assigned to differences in percentages of responses are based on Kelley's formula in which the correlation between responses determines in large

TABLE 23

ILLUSTRATING THE SELECTION OF ITEMS FOR BS SCALES

Item	Group	Percentage			Weights			*d* Score on Internist Scale	
		Like	Indifferent	Dislike	Like	Indifferent	Dislike		
No. 4								I Scale	
	P_{med}	21	30	49				19	
	I	16	23	61	0	−1	1	38	
	S	16	30	54				24	
	Pa	19	32	49				17	
	Ps	23	29	48				19	
No. 10								I, S & Ps Scales	Pa Scale
	P_{med}	52	31	17				56	39
	I	70	25	5	2	−1	−1	110	105
	S	71	22	7	2	−1	−1	113	106
	Pa	75	21	4	2	−1	−2	125	121
	Ps	77	18	5	2	−1	−1	131	126
No. 6								I, Pa Ps Scales	
	P_{med}	36	39	25				11	
	I	45	38	17	1	0	−1	28	
	S	37	38	25				12	
	Pa	48	40	12	1	0	−1	36	
	Ps	48	3[illegible]	14	1	0	−1	34	
No. 17								I Scale	S Scale
	P_{med}	28	42	30				−28	−2
	I	22	45	33	−1	0	0	−22	−11
	S	36	42	22	1	0	−1	−36	14
	Pa	28	40	32				−28	−4
	Ps	28	44	28				−28	0

P_{med} = Physicians-in-general reference group
I = Internists
S = Surgeons
Pa = Pathologists
Ps = Psychiatrists

part the weight (4, 5).

With item No. 10 we have a case where the item was used on all four BO scales. Weights of 2, −1, and −1 were assigned responses of liking, indifference, and disliking, respectively, on the Internist, Surgeon, and Psychiatrist Scales, and weights of 2, −1, and −2 on the Pathologist Scale. Once the four weights are brought together it is evident that use of this item on all four scales will produce a high correlation among them.

To explain the revision it is necessary at this point to define *d* scores. Given an item with *n* responses to each of which are attached scoring weights, then a man's score on the item is $W_1 R_1 + W_2 R_2 + W_3 R_3 +$ etc. where W equals the weight and R equals the man's response (viz., "1" if he marks that response, and "0" if he does not). The sum of such totals for all the items in the test is the man's raw score. So far the procedure is exactly that employed in scoring the man's blank on the scale.

Similarly, the mean score of *n* persons on a single item is

$$1/N\,[\,(W_1 \times \text{number replying}) + (W_2 \times \text{number replying}) + (\text{etc.})]\text{.}$$

Thus, if 80 physicians like "surgeon," 14 are indifferent, and 6 dislike the activity and the weights are, respectively, 4, −1 and −4, then the mean score on the item is

$$1/100\,[(4 \times 80) + (-1 \times 14) + (-4 \times 6)] \text{ or } 2.82.$$

The sum of such totals for all the items in the test is the mean raw score of the group.

By this procedure the mean score of a group can be obtained without scoring the individual blanks on the appropriate scale. All that is needed are the weights for the different responses to each item and the tally of responses of the group (6).

The d score is the *total, not mean, score of 100 persons on a single item*. In the above example it would be 282, not 2.82. Such d scores indicate the contribution of each item for each of the four specialist groups on each of the four scales. Table 23 shows the contribution of item No. 4 on the Internist Scale to be 38 for internists, 24 for surgeons, 19 each for P_{med} and psychiatrists, and 17 for pathologists. This is a poor item, approaching the lower limit of acceptability.

Three standards were set up in terms of which items on the BO scales were accepted for the BS scales. First, an item was rejected if higher d scores were obtained by any specialty group than the group for which the scale was designed. Second, an item was retained only if its d score was greater by 6 or more than the average of the d scores on the other three scales. It is not necessary to consider the P_{med} d scores since such scores cannot exceed the d scores for a group on its own scale, so long as the weights are properly assigned. Since the revision was undertaken to increase the differentiation between internists and psychiatrists, and between internists and pathologists—particularly with the Internist Scale—a third standard was maintained, namely, to accept items that discriminated in these directions even if they did not discriminate between internists and surgeons.

Consider now the application of these three standards to item No. 10 in Table 23. The weights for Internist, Surgeon, and Psychiatrist BO Scales are 2, −1, and −1. The psychiatrist d score is 131, which is higher than the d scores for the other groups and also the 131 exceeds the average of the other three d scores, i.e., 115, by 16. The item is accepted for the Psychiatrist Scale and rejected for the Internist and Surgeon Scales. The weighting is different for the Pathologist Scale, but pathologists have a lower d score than psychiatrists on that weighting; conseq[illegible] the item is rejected for the Pathologist Scale.

As far as the first two standards go, item No. 6 should be used in the Pathologist Scale since the pathologists have a higher d score than the other three groups and its d score is more than six greater than the average of the other three d scores. But it discriminated very little between internists and either pathologists or psychiatrists where greater discrimination was desired, while it discriminated considerably between pathologists and surgeons where the discrimination was already ample. The item was accordingly rejected.

Item No. 17 is an example of an item that can be used successfully on two different scales. This is because the weights are reversed on the Internist and Surgeon Scales. The item contributes very little to the Internist Scale, barely meeting our requirements, but it makes a good contribution to the Surgeon Scale.

TABLE 24

DIFFERENTIATION OF INTERNISTS FROM OTHER GROUPS ON THREE INTERNIST SCALES
(Vocational Interest Blank Only)

Group	I_a (BS) Scale		I_b Scale		I_c Scale		Per Cent Overlapping with Internists		
	Raw Mean	*SD*	Raw Mean	*SD*	Raw Mean	*SD*	I_a Scale	I_b Scale	I_c Scale
Internist	13.5	8.92	6.7	7.3	22.86	12.94	—	—	—
Surgeon	1.12	10.84	1.8	6.6	19.4	10.7	53	73	88
Pathologist	6.91	11.04	4.16	7.0	24.06	11.68	74	86	−96*
Psychiatrist	7.14	10.94	7.78	6.18	27.36	11.48	75	−94	−85
P_{med}	1.62	11.82	−.96	7.74	9.52	12.36	57	61	60

* Minus sign indicates the group scores higher than internists.
I_a Scale = 65 items selected for BS Internist Scale.
I_b Scale = 21 items which approximate standards for BS Scale.
I_c Scale = 96 items which are definitely below BS standard.

It might be pointed out that negative *d* scores are as useful as positive *d* scores provided they meet the two standard requirements.

Table 24 illustrates what happens when items are selected on the above basis for the Internist Scale on the Vocational Interest Blank. There were 182 items on the Internist BO Scale on the Vocational Interest Blank. Only 65 were selected for the BS Scale. Of the remainder, 21 items closely approximated the standards set up for selection and 96 were definitely rejected. These three groups of items were scored separately as though they were separate scales and are referred to as scales Ia, Ib, and Ic, respectively. The Ia scale is markedly better than the Ib scale, and both are clearly superior to the Ic scale. On this third scale both pathologists and psychiatrists score higher than internists. The superiority of the Ia scale over the other two scales is shown also in terms of overlapping. Internists and psychiatrists overlap only 75 per cent on scale Ia and overlap −94 and −85 per cent on scales Ib and Ic, respectively. The minus signs indicate that the overlapping is in the wrong direction; namely, that psychiatrists score higher than internists on the Internist Scale.[2]

5. *Evaluation of the BS Scales*

Comparisons of the effectiveness of the various BO and BS scales are given in Table 21. When both blanks are used, the average overlapping of specialist groups with one another is decreased from 52 to 42 per cent, which represents an increase in the average biserial correlation from .67 to .77. The poorest differentiation, represented by an overlap of 61 per cent and an r_{bis} of .56, is still fairly good.

It may help to understand these results to note that physicians and engineers overlap 42 per cent on the Engineer Scale on the Vocational Interest Blank. The medical specialist scales differentiate, on the average, one specialist group from another and from P_{med} to the same degree that physicians are differentiated from engineers.

The BS scales are a real improvement from the standpoint of validity, especially in differentiating internists from psychiatrists and pathologists. But they contain far fewer items and, presumably as a result, they have lower reliability.

---

[2] See footnote 1, Technical Note B.

If the number of items is plotted against the reliability, two things are evident. First, there is an almost direct relationship between the number of items and the odd-even estimates of reliability, if data from the BO and BS scales are considered separately. Second, the BS scales achieve a higher reliability per number of items used than do the BO scales. Approximately 300 items are needed for a reliability of .89 on the BO scales in contrast to 225 items on the BS scales. Evidently, the selected items have higher reliability than the rejected items.

The subject of reliability is discussed in detail in Technical Note F. It was decided to use the BS scales rather than the BO scales on the ground that validity is more important than reliability.

6. *Limitations of Selected-Item Procedure*

Where the objective is to differentiate subgroups from the total and at the same time subgroups from one another, the selected-item procedure is superior to the usual method of selecting interest items. The procedure has been successful when four medical specialist subgroups have been contrasted. It seems at this time that such a procedure could not be employed if as many as ten subgroups were to be so differentiated. Theoretically it could be done, as far as we know; but from the practical point of view, it is very doubtful.

With the selected-item procedure, an item involving two possible responses, such as the forced-choice items of the Medical Specialists Preference Blank, can be used on only two scales; and an item involving three possible responses can be used on only three scales. On the typical like-indifferent-dislike item, weights such as −1, 1, 0; −2, 0, 2; and 1, −1, −1 could be used on three different scales. If all the 571 items on both blanks were used to theoretical capacity, this would mean that the average item would be used on 2.79 scales. But actually the items were used, on the average, on only 1.13 scales.

Assuming that 225 items are needed to give a reliability of .89, it means that 199 items are actually needed per scale, as each item is used on 1.13 scales. For four scales 796 items would be needed, 225 more than on the present two blanks. For 10 scales 1990 items would be necessary, which would require 1419 new items.

It is doubtful if 1419 additional good items [illegible]ld be found without great difficul[illegible], of course, to be expected that some of the items not used on the present four scales would be used on the six new scales and some of the present items not used to capacity could also be used more fully. Nevertheless, the number of new items necessary for ten scales presents an almost impossible task. Another objection to using a test of 1990 items is that very few would voluntarily fill out such a blank, so that it would be difficult to obtain the necessary data from criterion groups to develop the scales.

A more feasible procedure would be to determine the relationship of the other medical specialist groups to the present scales. It is likely that a number of specialist groups will be found closely allied to one of these four groups. Thus, urologists and orthopedic surgeons tend to score like surgeons and not like internists, pathologists, or psychiatrists. The present scales may thus be shown to predict interest in the four specialties and also in a considerable number of the remaining specialties.

Once it is established that certain specialists are primarily like surgeons as distinct from internists, pathologists, and psychiatrists, it might be possible to establish scales for these specialist groups, not in terms of P_{med} but in terms of a point of reference composed entirely of surgeons. Medical students could be scored on the present four BS scales plus additional scales which would indicate whether their interests were, for example, like surgeons, or like urologists, orthopedic surgeons, or neurological surgeons.

TECHNICAL NOTE C

THE SPECIALIZATION LEVEL SCALE (5)

By Milton G. Holmen, Ph.D.

THE Specialization Level Scale was constructed to help the young doctor answer the question, "Should I specialize?" The first step in developing this scale was to average the responses of the four specialist groups (internists, pathologists, psychiatrists, and surgeons) to each possible response on the Vocational Interest Blank and the Medical Specialists Preference Blank. The responses of each specialist group were given equal weight in computing the averages used as the responses of the medical specialist group. Then the responses of the specialists were compared with those of the P_{med} (physicians in general) reference group. Weights were assigned as described in Technical Note B for the BO scales.

The four specialist groups differ somewhat from each other. It was believed that averaging the responses of the four groups would eliminate any group peculiarities. It was thought, also, that subtracting the responses of physicians in general from the averages of the four specialist groups would eliminate medical interests as such from the scale and leave only the interests of specialists. These interests of specialists should, thus, be relatively independent of the field of specialization.

TABLE 25

ILLUSTRATION OF METHOD OF DETERMINING WEIGHTS FOR AN ITEM ON THE SPECIALIZATION LEVEL SCALE

Groups	Per Cent Responding		
	Like	Indifferent	Dislike
Surgeons	65	28	7
Internists	50	35	15
Psychiatrists	55	29	16
Pathologists	82	16	2
Mean of Specialists	63	27	10
Physicians in General	51	36	13
Difference	12	−9	−3
Weight assigned	1	−1	0

A Specialization Level Scale was developed for each of the two blanks by performing the computations illustrated in Table 25, for each of the items on the Vocational Interest Blank and the Medical Specialists Preference Blank. The Specialization Level Scale on the Vocational Interest blank contains 240 weights on 140 items. The scale on the Medical Spe[illegible] Preference Blank has 94 weig[illegible] 71 items.

The reliability (split-half corrected by the Spearman-Brown formula) of the scale using both blanks is .79. The Vocational Interest Blank part of the scale has a reliability of .77, and the Medical Specialists Preference Blank part of it has a reliability of .47. The blanks used for obtaining this reliability were those of 100 physicians in general. A test-retest procedure would have been preferable as an estimate of reliability, but the practical problems involved in getting test-retest scores on this group more than offset the advantages to be obtained by their use. The reliability of only .47 for the Medical Specialists Preference Blank part of the scale obviously precludes use of this blank by itself. The reliabilities of .77 and .79 are lower than one might desire, but not so low as to preclude use of the scale, if this fact is kept in mind when interpreting the scores.

In developing norms for the scale, the

TABLE 26

STATISTICAL CHARACTERISTICS OF DISTRIBUTIONS OF SCORES ON THE SPECIALIZATION LEVEL SCALE

Group and Measure	Vocational Interest Blank	Medical Specialists Preference Blank	Both Blanks
Medical Specialists			
Mean	50	50	50
SD	10	10	10
Physicians in general			
Mean	38.33	43.36	39.03
SD	10.20	9.65	9.85
Difference between means			
Mean difference	11.67	6.64	10.97
SD of Diff. between Means	1.14	1.09	1.10
Critical Ratio of Diff.	10.24	6.09	9.97

Correlation between scores on the two blanks = .47

blanks of 100 specialists in each of the four fields were scored on the scale. The scores of this group ($N = 400$) were then transmuted into standard scores so that the distribution had a mean of [illegible] and a standard deviation of 10.

The standard scores on the Specialization Level Scale were then obtained from the blanks of 100 physicians in general. These scores were compared with those of the specialists. The test used to determine the significance of the difference between the two distributions was the critical ratio obtained by dividing the difference between the means by the standard error of that difference. Table 26 presents these means, standard errors, and critical ratios. These data clearly indicate that the two groups are separated by the use of each blank separately and by the two blanks combined.

A more practical measure of the usefulness of this scale is how well it actually classifies members of various groups into the categories of "A," "B," etc. The distributions of the four groups of specialists and of three other medical groups are shown in Table 27.

Research on the Vocational Interest Blank section of the Specialization Level

TABLE 27

DISTRIBUTIONS OF RATINGS OF VARIOUS MEDICAL GROUPS ON THE SPECIALIZATION LEVEL SCALE

Group	Per Cent Obtaining Each Rating				
	A	B+	B	B−	C
Internists	69	16	9	5	1
Pathologists	68	19	8	3	2
Psychiatrists	86	8	4	1	1
Surgeons	52	24	10	12	2
Medical Specialists	69	17	8	5	1
Physicians in General*	28	18	22	14	18
Medical School Seniors	37	25	17	9	12
Army Command and Staff Medical Officers	37	22	16	13	12

* Since physicians in general were used in constructing the scale, their ratings are probably spuriously low. However, an average change of less than one standard score would be expected on another similar sample (5, p. 647).

Scale has yielded some interesting findings. Within three nonmedical subject-matter areas (social science, physical science, and accounting), occupational groups are ranked in the same order by their mean specialization-level score as by their mean educational-level. This may be taken to indicate that the scale measures liking for advanced education. To test whether this relationship would hold where the education was not of a specialized nature, scores of two groups of graduate students in business administration were obtained. These students had been enrolled in the Stanford Graduate School of Business working for the degree of Master of Business Administration. Half of them received this degree, and the other half dropped out or failed before completing the requirements for it. The Specialization Level Scale does not separate these two groups.

The Specialization Level Scale does separate chemists with Ph.D. degrees from those having M.S. degrees and those having B.S. degrees. The separation of chemists with M.S. degrees from those with B.S. degrees is not at a significant level (CR of .81), but is in the expected direction. Since advanced training in chemistry is highly specialized, and that in the Graduate School of Business is intended to be broadening, we may tentatively accept the hypothesis that what the scale measures is willingness to narrow one's vocational activities, as required of a specialist, rather than mere tolerance for advanced education.

TECHNICAL NOTE D

PHYSICIAN AND PSYCHIATRIST SCALES ON THE VOCATIONAL INTEREST BLANK

THE guidance of college men and potential college men has been assisted by the many interest scales of the Vocational Interest Blank for Men. These scales are based on the P_{men} (men in general) reference group which is representative of men in occupations normally entered by men with a college education and is thus not representative of all men.

This Technical Note discusses two scales for the Vocational Interest Blank. There has been a Physician Scale for many years which has helped answer the question, "Should I be a physician?" This scale has been revised as part of this research project. The original and revised physician scales are com[illegible]

The possible need for a psychiatrist scale in addition to the physician scale for use with college students will also be discussed. As compared to men in general, are the interests of psychiatrists substantially different from those of the average physician?

1. *Revision of Physician Scale*

The revised physician scale for the Vocational Interest Blank is based on the P_{med} (physicians in general) reference group. This sample of 500 physicians is representative of all physicians in the country. Surprisingly good agreement has been obtained between scores on the original scale and subsequent careers twenty years later (8). Will equally good or better predictions be obtained by use of the revised scale?

Scores on the original and revised scales correlate .85. In the revision of 28 other occupational scales, the average correlation between the original and revised scales was .93. In only two of the 28 revisions did the correlations fall below .85, namely, .83 and .81. Revision of these 28 scales involved an increase in size of the criterion group, on the average, of 64 cases (from 200 to 264 cases) and in most cases involved a revision of the men-in-general group as the point of reference. Revision of the physician scale involved an increase in size of the criterion group of 123 cases (from 377 to 500 cases) but no change in point of reference. In terms of these factors one would not expect as low a correlation as .85 between the original and revised physician scales. There was one other factor which may have contributed appreciably to this low correlation. The original scale was based very largely on graduates of Stanford Medical School, whereas the revised scale is based on graduates of all the medical schools in the country.

One measure of the goodness of an interest scale is the ratio of raw mean score divided by the raw standard deviation. On this basis the revised scale is greatly superior to the original scale—the ratios being, respectively, 2.72 and 1.84. If these ratios are multiplied by 10 and subtracted from 50, the result is the raw zero score expressed in standard scores, i.e., 22.7 and 31.6. The chance score always approximates the raw zero score, being 21.7 for the revised scale and 29.2 for the original scale. From the standpoint of chance there is less likelihood of obtaining a high score on the revised scale than on the original scale. Both scales are about equally reliable (.873 and .886).

TABLE 28

CORRELATIONS OF THE REVISED AND ORIGINAL PHYSICIANS SCALES WITH OTHER VOCATIONAL INTEREST BLANK SCALES
($N = 285$ Stanford Seniors)

Scale	Physician Scale (revised)	Physician Scale (original)
Psychologist (rev.)	.75	.58
Physician (original)	.85	—
Dentist	.83	.87
Carpenter	.34	.23
Personnel Manager	−.24	−.51
Office Worker	−.54	−.72
Occupational Level	−.27	.03
Masculinity—Femininity	.02	−.06
Osteopath	.80	.69
Internist A	.93	.81
Surgeon A	.95	.85
Pathologist A	.93	.79
Psychiatrist A	.79	.54

Correlations between scores on the original and revised Physician Scales with scores on several other scales are given in Table 28. The four medical specialist A scales[1] correlate substantially higher with the revised scale than with the original scale (average .90 and .75, respectively). Aside from this difference, the corresponding correlation coefficients differ relatively little.

It appears that the general relationships between the interests of physicians and other occupational groups as established by the original scale are maintained to a large extent by the revised scale. Since the revised scale is based on a more recent and representative sample of physicians than was the original scale, it is recommended that this revised Physician Scale be used in counseling college men.

2. *Need for a Psychiatrist Scale in Addition to a Physician Scale*

As measured by the Vocational Interest Blank, surgeons differ least of all in their interests from physicians in general, internists are second, pathologists third, and psychiatrists fourth (see Table 29). This relationship holds whether differences in mean scores or per cent of overlapping are considered, and holds also on both A and BO type scales.

The same relationships hold when correlations are considered between each of the four specialist A scales and the revised Physician Scale. The Surgeon Scale correlates .95 with the Physician Scale, the

TABLE 29

COMPARISONS OF MEDICAL SPECIALIST SCALES
(Based on Vocational Interest Blank Only)

Type of Scale	Scale	Groups Compared	Difference in Means	Per Cent Overlapping
A	Internist	I vs. P_{med}	9.0	67
	Surgeon	S vs. P_{med}	6.5	75
	Pathologist	Pa vs. P_{med}	11.85	60
	Psychiatrist	Ps vs. P_{med}	13.25	52
BO	Internist	I vs. P_{med}	14.9	48
	Surgeon	S vs. P_{med}	14.5	53
	Pathologist	Pa vs. P_{med}	17.6	43
	Psychiatrist	Ps vs. P_{med}	18.2	35

P_{med} = Physicians in general, $N = 100$.
I = Internists, $N = 100$.
S = Surgeons, $N = 100$.
Pa = Pathologists, $N = 100$.
Ps = Psychiatrists, $N = 100$.

---

[1] For definitions of A and BO type scales refer to Technical Note B.

TABLE 30

Rotated Factor Loadings of Vocational Interest Blank Scales
(Factors II and III rotated 51° clockwise around I)

Scale	Factors				L²
	I	II	III	IV	
1. Psychologist (rev.)	728	560	278	126	937
2. Physician (original)	897	−284	224	−244	995
3. Dentist	873	−304	−198	−298	982
4. Carpenter	447	−122	−811	228	924
5. Personnel Manager	−310	793	000	−078	732
6. Office Worker	−595	287	−505	−018	691
7. Occupational Level	−087	−400	815	−070	851
8. Masculinity-Femininity	142	−230	−550	188	410
9. Osteopath	741	061	−215	−584	940
10. Internist A	925	260	199	−076	969
11. Surgeon A	953	109	098	−268	991
12. Pathologist A	972	216	−054	159	1019
13. Psychiatrist A	708	659	127	−221	1000
14. Internist BO	127	506	695	190	790
15. Surgeon BO	187	190	374	−157	236
16. Pathologist BO	754	256	−038	420	812
17. Psychiatrist BO	−137	879	162	−125	835

Decimal points omitted.

Internist Scale .93, the Pathologist Scale .93, and the Psychiatrist Scale .79.

Correlations of .93 to .95 [illegible] express perfect agreement, but they give no warrant for supposing that better guidance of college men can be given through the use of Internist, Surgeon, and Pathologist Scales in addition to the use of the Physician Scale. The lower correlation of .79 between physicians and psychiatrists raises the question, Should college men be tested on both scales instead of on only the Physician Scale?

It appears that physicians and psychiatrists differ more than physicians and dentists (.83), and about the same as physicians and osteopaths (.80), physicians and artists (.79), and physicians and physicists (.78) (5). On this basis there is warrant for the use of a psychiatrist scale in addition to a physician scale. Evidence from factor analysis distinctly supports this conclusion.

In order to obtain a better conception of the interrelations between the A and BO scales, a factor analysis was made based on the intercorrelations of scores on 17 scales. Included in the matrix were 7 occupational scales, the 4 medical specialist A type scales, the 4 medical specialist BO type scales, the Occupational Level Scale, and the Masculinity-Femininity Scale. The scores are based on the Vocational Interest Blank only.

Several rotations of factors were made leading to the conclusion that the factor loadings as shown in Table 30 would be most useful. In this table the original factor loadings of Factors II and III were rotated 51° clockwise around Factor I.

The first factor may be named "medical interest." The scales and loadings expressive of this interest are:

.972	Pathologist A
.953	Surgeon A
.925	Internist A
.897	Physician (old)
.873	Dentist
.754	Pathologist BO
.741	Osteopath
.728	Psychologist (rev.)

.708	Psychiatrist A
−.595	Office Worker

Psychiatrists have a lower loading than any other medical group, even slightly lower than osteopaths and psychologists on the A scales. They score lower than the other three specialist groups on the BO scales.

Theoretically, medical interest should be eliminated in the BO scales since similarities in interests between physicians in general and the four specialist groups are ignored in the weighting of items. This holds true as shown by the drop in average loading for Factor I from .862 on the A scales to .059 on the Internist, Surgeon, and Psychiatrist BO Scales. For some unknown reason, the Pathologist BO Scale has the high loading of .754 in Factor I.

The second factor may be named "interest in people." The scales and loadings expressive of this interest are:

.879	Psychiatrist BO
.793	Personnel Manager
.659	Psychiatrist A
.560	Psychologist (rev.)
.506	Internist BO

Psychiatrists score much higher than other medical groups on this factor, so that as far as this factor goes they are to be grouped with personnel men and psychologists rather than with medical men. No interpretation is made of Factors III and IV. It appears that psychiatrists have less "medical interests" and more "interest in people" than the other three medical specialists and physicians in general.

These results suggest that a Psychiatrist Scale, in addition to the Physicians Scale, might be useful in counseling college men. There are indications that the propor[illegible] of physicians doing psychiatric[illegible]s increasing substantially. It may [illegible]at a certain number of men shoul[illegible]nter the medical schools whose Physic[illegible] Scale scores are only fair but whose Psychiatrist Scale scores are high.

Technical Note E

DEVELOPMENT OF THE MEDICAL SPECIALISTS PREFERENCE BLANK

One of the first decisions made on this project was to develop a blank specifically for use by the medical profession. All possible types of items that could be conceived of for this type of blank were freely discussed among the members of the project. A substantial number of conferences was held with interested members of the medical profession, and the ideas of some physicians and psychologists were obtained by correspondence.

One early decision was to emphasize the use of the "forced-choice" type of item which does not permit of an "indifferent" response. It was believed that this type of item would be more effective than the "Like-Indifferent-Dislike" item in forcing fine discriminations among quite similar professional groups. Later research by Zuckerman did not support this belief (10).

Another decision was to assign primary responsibility for the development of items in different areas to each of the four graduate assistants working on the project. They were Robert E. Adamson, Ralph J. Garry, Milton G. Holmen, and John V. Zuckerman. The several types of items that were constructed will each be considered by indicating: (1) kind of item, (2) responsible assistant, (3) number of such items constructed, (4) brief statement of rationale, and (5) sample item.

1. Personal and Professional Relationships–Holmen–80 items. The several types of specialists would differ as to their expressed preferences regarding the characteristics of their professional associates. Sample item—Which characteristic do you consider more desirable in an associate? (A) Conducts office in very businesslike manner. (B) Conducts office in very friendly, informal manner.

2. Medical and Premedical School Subjects–Holmen–45 items. There would be stable differences among specialists as to liking or disliking certain subjects while in school. Sample item—Indicate whether you liked, were indifferent to, or disliked, Comparative Physiology, Cytology, etc.

3. Casual Acquaintances and Dinner Companions–Holmen–40 items. Specialists would tend to differ in preferring the company of individuals in different occupations. Sample items—On a cross-country train trip, would you like or dislike to make the acquaintance of a: Banker, Artist, Reporter, etc.? At a large dinner party, would you like or dislike to be seated beside a woman who was a: Club Leader, Tennis Professional, Librarian, etc.?

4. Magazines–Holmen–20 items. There would be differences in preferences for magazines. Sample item—Do you like each of the following magazines: *Life, Colliers, Reader's Digest,* etc.?

5. Amusements – Holmen – 20 items. There would be differences in preferences for amusements. Sample item—Indicate which of each of the following pairs of amusements you prefer: Play Bridge or Play Poker; Western Movies or Whodunit Movies, etc.

6. Occupational Preferences–Zuckerman–90 items. It was postulated that there are four modes of dealing with the environment, namely, analytic, visual, social, and manipulative, which are measured by descriptions of occupational and avocational activities. It was believed that

the different medical specialists would prefer different modes of behavior. Sample item—Indicate which of each pair of occupational activities you would prefer: (A) Interpret spoken passages for individuals who speak only foreign languages. (B) Install aerials for radio receiving sets; set up poles and attach the wires to them.

7. Recreational Activities–Zuckerman –78 items. Essentially the same rationale as for the Occupational Preference items. Sample item—Indicate which of the activities you prefer: (A) Build a home-made radio set. (B) Look at a picture magazine.

8. Biographical Data – Adamson – 86 items. It was believed that differences in background and early behavior would differentiate among the several types of specialists. Sample item—In grade school, did you (A) Get into many fights? (B) Avoid fights by trying to "talk it over"? (C) Fight only when forced to?

9. Type of Leadership–Garry–86 items. Three types of leadership were postulated which could be crudely labeled autocratic (leader-centered), democratic (group-centered), and laissez-faire (individual-centered). It was believed that surgeons would tend to prefer autocratic leadership, psychiatrists the democratic, and internists the laissez-faire. Sample item—Do you prefer the leader who: (A) Assumes that group members know routine procedures, or (B) Gives specific orders in routine tasks.

Experimental blanks containing these items were administered to approximately 300 Medical Corps officers at Letterman, Fitzsimons, and Brooke Army Hospitals. Included in this group were the interns, and the resident and staff doctors of the Medical, Surgical, Neuropsychiatric, and Laboratory Services. Item counts were made of the responses of the various groups.

The best items of these experimental blanks were selected for use in Form C of the Medical Specialists Preference Blank, which included 194 items. In selecting items for this blank, consideration was given to the ability to differentiate among the several groups of Army Medical Corps Officers. However, the selection was not done on a strictly objective basis.

Form C of the Medical Specialists Preference Blank was used in the mailing to the large civilian groups of Internists, Surgeons, Pathologists, and Psychiatrists, and to the physicians-in-general group selected by the American Medical Association. Item counts on these groups were the b[illegible] for the development of the sca[illegible] this research.

Fo[illegible] represents a revision of this blank, which retained all items that were weighted on the scales. The section on Types of Leadership was eliminated because of generally poor differentiation and because of the large number of items not answered. In order to facilitate scoring, the four- and five-choice Biographical Information items were restructured as three-choice items. This form contains 171 items in the following parts: Professional Associate–50 items; Biographical Data–24 items; Medical School Subjects –35 items; Traveling–15 items; Occupations–27 items; and Activities–20 items.

Form D was used in obtaining data on the following groups: medical school seniors, Army interns, Army Command and Staff medical officers, orthopedic surgeons, urologists, neurological surgeons, and pediatricians.

TECHNICAL NOTE F

CROSS VALIDATION AND RELIABILITY

THE true validity of the medical specialist scales, that is, their ability to predict future behavior, can only be determined by follow-up studies extending over a period of several years. Arrangements have been made for such studies by collecting blanks from approximately 750 medical school seniors.

In the meantime, any cross-validation data are of interest in estimating the probable shrinkage in effectiveness when administered to new groups. Unfortunately, most of the available criterion populations were used in establishing these scales. However, some blanks were received after the item counts were made and these blanks constitute [illegible] source of cross-validation data. The[illegible] are small, ranging in size from [illegible].

Table 31 compares the distributions of scores on the blanks used in the item counts on which the scales were based with the scores on the blanks received later. No change is indicated on the Internist and Pathologist Scales and the changes on the other two scales are of doubtful significance. It appears that a few blanks in the 30-34 standard score range by the surgeons and psychiatrists may largely account for the changes in means on these scales.

Another method of examining the cross-validation groups is to determine the percentage of overlap of the various pairs of groups on the four medical-specialist scales. Table 32 presents the results of such an analysis. It is interesting that five out of the six comparisons in-

TABLE 31

DISTRIBUTIONS OF SCORES OF CRITERION GROUPS AND SMALL CROSS-VALIDATION GROUPS ON THE MEDICAL SPECIALIST SCALES

Standard Score	Per Cent of Internists on Internist Scale		Per Cent of Surgeons on Surgeon Scale		Per Cent of Pathologists on Pathologist Scale		Per Cent of Psychiatrists on Psychiatrist Scale	
	Item Count Blanks $N=211$*	Extra Blanks $N=80$	Item Count Blanks $N=188$*	Extra Blanks $N=58$	Item Count Blanks $N=100$*	Extra Blanks $N=24$	Item Count Blanks $N=167$*	Extra Blanks $N=35$
65+	7.1	6.2	7.4	5.2	6.0	8.3	6.6	5.7
60–64	10.0	8.8	9.6	5.2	10.0	8.3	11.4	2.9
55–59	14.7	20.0	17.0	10.3	16.0	29.2	18.0	14.4
50–54	17.1	18.8	21.3	20.7	21.0	12.5	22.2	22.8
45–49	19.8	13.8	21.3	22.4	15.0	4.3	15.6	22.8
40–44	18.5	16.2	14.3	15.5	16.0	20.8	11.4	11.4
35–39	8.5	8.8	5.3	5.2	10.0	8.3	9.6	5.7
30–34	1.9	1.2	2.7	13.8	4.0	8.3	1.8	8.6
29–	2.4	6.2	1.1	1.7	2.0	—	3.4	5.7
Total	100%	100%	100%	100%	100%	100%	100%	100%
Mean	49.70	49.40	50.85	47.15	49.55	50.35	50.40	47.55
SD	9.45	10.15	8.80	9.80	9.60	10.25	9.75	10.00
CR of Diff.		.29		2.57		−.35		1.54

* The item counts used in constructing the scales were based on samples of approximately 400, as shown in Table 6. The smaller number shown here represents the doctors who returned both the Vocational Interest Blank and the Medical Specialists Preference Blank.

TABLE 32

COMPARISONS OF THE OVERLAP OF CRITERION GROUPS AND SMALL CROSS-VALIDATION GROUPS ON THE MEDICAL SPECIALIST SCALES

Scale	Groups Compared	Per Cent of Overlap	
		Criterion Groups	Cross-Validation Groups*
Internist	Internists vs. Surgeons	31	42
	Internists vs. Pathologists	61	58
	Internists vs. Psychiatrists	60	60
Surgeon	Surgeons vs. Internists	24	37
	Surgeons vs. Pathologists	31	48
	Surgeons vs. Psychiatrists	18	31
Pathologist	Pathologists vs. Internists	53	51
	Pathologists vs. Surgeons	44	46
	Pathologists vs. Psychiatrists	42	38
Psychiatrist	Psychiatrists vs. Internists	57	62
	Psychiatrists vs. Surgeons	38	51
	Psychiatrists vs. Pathologists	46	47
	Average	42%	48%

* Size of cross-validation groups—Internists $N = 80$, Surgeons $N = 58$, Pathologists $N = 24$, Psychiatrists $N = 35$.

TABLE 33

ESTIMATES OF THE RELIABILIT[illegible] ALES

(Correlations of Odd-Eve[illegible]
$N = 100$ Physicians in g[illegible]

Scale	Type	Vocational Interest Blank			Medical Specialists Preference Blank			Both Blanks		
		Items	Raw	Corrected	Items	Raw	Corrected	Items	Raw	Corrected
Internist	BS	65	.47	.64	32	.39	.56	97	.52	.69
	BO	182	.62	.76	96	.70	.83	278	.75	.86
Surgeon	BS	87	.58	.73	51	.46	.63	138	.65	.78
	BO	184	.60	.75	85	.46	.63	269	.65	.79
Pathologist	BS	116	.67	.80	59	.55	.71	175	.67	.80
	BO	229	.72	.83	103	.47	.64	232	.74	.85
Psychiatrist	BS	165	.77	.87	71	.55	.71	236	.80	.89
	BO	250	.77	.87	110	.64	.78	360	.82	.90
Specialization Level	BO	149	.62	.77	71	.31	.47	220	.65	.79
Surg—Int	C	194	.65	.79	102	.67	.80	296	.77	.87
Path—Int	C	186	.70	.82	91	.49	.65	277	.69	.82
Psych—Int	C	230	.76	.86	102	.20	.34	332	.69	.82
Path—Surg	C	240	.68	.81	95	.56	.72	335	.73	.84
Psych—Surg	C	253	.77	.87	120	.68	.81	373	.85	.92
Psych—Path	C	252	.75	.86	121	.52	.69	373	.79	.88

Items = Number of items weighted on the scale.
Raw = Product moment correlation between score on odd versus even items.
Corrected = Raw corrected by Spearman-Brown formula.

volving surgeons show a marked increase in overlap while all the other comparisons show little difference between the criterion and cross-validation groups. In fact, the six comparisons which do not involve surgeons have the same average overlap in each group. It is not possible, at this time, to determine whether these results with the surgeons are an artifact due to a few peculiar blanks in the small cross-validation group or because the Surgeon Scale is less stable than the others.

The validity of scales, that is, how well the various scales differentiate between groups, has been emphasized in this research report. Another measure of the goodness of a scale is reliability. What is the reliability of the various scales?

One estimate of reliability is the correlation between scores on even-numbered items and scores on [illegible]-numbered items on each scale. [illegible] type of reliability refers primarily [illegible] internal consistency of the scale. Table 33 reports the raw odd-even correlations for each scale, as well as these correlations corrected by the Spearman-Brown formula. The number of items on each scale is also indicated.

The scales which are of primary interest are the Internist, Surgeon, Pathologist, and Psychiatrist BS scales. These four BS scales have an average reliability of .791 and contain, on the average, 162 items per scale. The corresponding BO scales have an average reliability of .850 and the average number of items per scale is 285.

Since there is a direct relationship, as based on our data, between reliability coefficients and number of items in the scales, the reduction in reliability from BO to BS scales is to be attributed to the decrease in number of items used in BS scales as compared to BO scales. This is particularly true with respect to the BS Internist Scale.

Another measure of reliability, referred to as the coefficient of stability, is based upon the correlation between test and retest scores, where the two blanks are completed within a relatively short period of time. Burnham reported such coefficients for eight interest scales for an interval of one week (2). These eight coefficients averaged .906 in contrast to .866 for coefficients based on odd-even comparisons. It is likely that the coefficients of reliability given in Table 31 are lower than they would be if based on test-retest scores.

The size of a correlation coefficient is affected by the range of scores on which the computations are based. The coefficients in Table 33 are based on blanks from the physician-in-general group. It seemed appropriate to use this group since the BO and BS scales were designed to differentiate physicians. The physician-in-general group is, however, a relatively homogeneous group.

The issue is whether or not scales with poorer validity and better reliability should be preferred to scales with better validity and poorer reliability (see Tables 21 and 33). BS scales have been selected over BO scales on the basis of greater validity. It is believed that validity is more important than reliability, that validity automatically necessitates reliability, and that the measures of internal consistency in Table 33 are not complete measures of reliability.

The answer to the above question must wait until there can be a follow-up of medical students which will ascertain the extent to which scores of medical students on the two types of scales are related to subsequent medical careers.

APPENDIX I

CHARACTERISTICS OF SPECIALIST CRITERION GROUPS

A. Age In Years

Item		Internists	Surgeons	Pathologists	Psychiatrists	Physicians in General
Mean Range		36.9 32–39	38.4 31–44	40.5 31–49	39.3 32–46	40.9 24–55

B. Per Cent Residing in Each Geographical Area*

	Of All Physicians in U.S.	Internists	Surgeons	Pathologists	Psychiatrists	Physicians in General
1. New England	7.9	7.6	9.2	9.1	9.8	8.4
2. North Atlantic	29.6	26.8	26.0	31.7	39.9	30.4
3. Midwest	28.1	26.8	24.6	25.3	25.1	28.4
4. South	15.3	16.8	16.9	13.0	9.2	14.5
5. Southwest	5.8	6.2	6.[illegible]	4.4	3.3	5.4
6. Mountain	2.3	2.6	[illegible]	3.5	2.4	2.3
7. Pacific	11.0	13.2	[illegible]	3.0	10.3	10.6
Total	100%	100%	10[illegible]	[illegible]00%	100%	100%

C. Per Cent Graduates of Medical Schools Located In Each Geographical Area*

Area		Internists	Surgeons	Pathologists	Psychiatrists	Physicians in General
1. New England		10.7	15.5	10.8	10.5	8.0
2. North Atlantic		31.1	32.1	36.6	36.6	30.8
3. Midwest		35.1	30.6	32.4	37.1	35.2
4. South		13.1	11.7	9.5	7.1	17.0
5. Southwest		3.8	3.3	1.7	2.8	4.2
6. Mountain		1.0	.6	1.5	1.3	.8
7. Pacific		5.2	6.2	7.5	4.6	4.0
Total		100%	100%	100%	100%	100%

---

* New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut. North Atlantic: New York, Pennsylvania, New Jersey, Delaware, Maryland, District of Columbia. Midwest: Ohio, Michigan, Indiana, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Kansas, Nebraska, Illinois. South: Virginia, West Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana. Southwest: Texas, Oklahoma, New Mexico, Arizona. Mountain: Idaho, Montana, Wyoming, Nevada, Utah, Colorado. Pacific: Washington, Oregon, California.

APPENDIX II*

REPORT OF MEDICAL INTERESTS

Purpose

The purpose of this report is to assist you in planning your career. It is based on your responses on the Vocational Interest Blank for Men[1] and the Medical Specialists Preference Blank.[2] The Vocational Interest Blank has been widely used for many years in assisting young men in making decisions regarding their vocational careers. The Medical Specialists Preference Blank has been developed specifically for use by the medical profession.

The extent to which your interests agree or disagree with those of successful men in certain types of work are indicated. Extensive research over the past twenty-five years shows that such measures have considerable accuracy in predicting whether one would like certain occupations or not. Continuing high motivation and job satisfaction appear to be related to these measures of common interests.

This is *not* a report of your *aptitudes* or *abilities* to do certain types of work. For exam[illegible] questions regarding your ability to get throu[illegible] school or to learn to perform surgical ope[illegible] answered by this report.

The measures of the interests [illegible]ns and of the four major medical specialti[illegible]een developed by the Medical Specialists Research Project, Stanford University.[3] The physician scale is based on the responses from a random sample of practicing M.D.'s selected by the American Medical Association in 1949 from their records. The specialists scales are based on the blanks of approximately 1800 diplomates of the American Boards of Internal Medicine, Surgery, Pathology, and Psychiatry and Neurology.

Explanation of Scores

Four questions confront anyone considering medicine as a career. First, should one be a physician or engage in some other occupation? Second, if a physician, should one specialize in one of the 32 recognized specialities to the extent of taking several years residency training and passing the board examinations? Third, if one is to specialize, which specialty should one select? And fourth, what types of activities within the medical profession would one like to do?

Test No. 1 answers the first question. Your interests are compared with the interests of the average doctor and the score indicates whether your interests are more in agreement with the interests of the average doctor or more in agreement with the average man engaged in those occupations which college men expect to enter. Table 1 gives the percentages of various groups who have ratings from A to C on the physician scale. If you have a B rating, for example, in physician interest it means that at least 84 percent of physicians have interests more akin to the average doctor than you do and at least 8 percent of physicians have interests less akin to the average doctor, than you do. Having a B— or C rating does not mean you cannot be a doctor, but that you probably will not enjoy treating patients and that there may be some other occupational activity which is more akin to your interests.

TABLE 1. RATINGS ON THE PHYSICIAN INTEREST SCALE

Rating	Percent of Practicing Physicians	Percent of Seniors in Medical School	Percent of Non-Medical College Students	Percent of Graduate Students in Business Administration
A	71	73	11	1
B+	13	14	10	2
B	8	8	13	4
B—	5	4	13	10
C	3	1	53	83
Total	100%	100%	100%	100%

---

[1] Vocational Interest Blank for Men (revised) Form M by Edward K. Strong, Jr., Stanford University.

[2] Medical Specialists Preference Blank, Form D (Research edition), by E. K. Strong, Jr., A. C. Tucker, R. E. Adamson, M. G. Holmen, J. V. Zuckerman, Stanford University.

[3] Contract No. W-49-007-MD-483 with the Surgeon General, U.S. Army.

* This is a reproduction of the form used in reporting medical interest scores.

The second question is answered by Test No. 2, i.e., do you have the interests of medical specialists or the interests of the average non-specialist doctor? About fifteen percent of physicians are today diplomates in some specialty. It is probable that this percentage will be doubled in the next twenty years. You may or may not wish to go through the long specialized training perfecting yourself in a particular specialty and may or may not desire to limit your practice as is required of a diplomate.

The third question is partially answered by the four Tests 3, 4, 5 and 6. At the present time tests are available for only the four specialities of Internist, Surgeon, Pathologist, and Psychiatrist. How you score on these four specialties may also give some idea of how you might score on other specialties which are akin to these four. For example, the surgeon's score should represent the areas of urology and orthopedic surgery quite well. Similarly cardiology and dermatology appear closely related to internal medicine.

A high score on one of the specialties coupled with a low score on Test 2 suggests that you will enjoy work in that activity but that you may not want to spend several years in rigorous preparation or to devote all your time to it.

The four Tests 3, 4, 5 and 6 apply only to men who have the interests of doctors. These scores are meaningless if you rate C or B— on the physician scale and questionable if you have a B rating.

Some men score about equally high on two, or even three, of the four specialties. In such cases information is given in No. 7 as to whether there is a real difference (yes), a probable difference, or the scores are tied.

The fourth question refers to types of work within the medical profession which will be interesting to you. The scores for the ten occupational scales listed in No. 8 may be helpful. Among graduates of medical schools some devote relatively little or no time to the treatment of patients, spending their time in a variety of activities, such as teaching, research, hospital administration, preventive medicine, marketing medical supplies, etc.

Your scores in No. 8 may be interpreted in two ways. First, with reference to how men score who are engaged in these occupations. An A rating means "yes," you have the interests of men in that occupation. A rating of B+ means "probably," etc. Such ratings are interpreted in the same way as your score in physician interest in No. 1.

For some purposes it is well to consider how your interests compare, not with men engaged in the ten occupations, but with doctors alone. The shaded areas provide the basis for such interpretations of your scores. Twenty-five percent of physicians score above the shaded area, fifty percent score within the shaded area, and twenty-five percent below this area. For example, if your score on office worker is a B, it means you *possibly* have the interests of office workers; but you have more of the interests of office workers than the great majority of physicians. Such a score [illegible] that you may like administrative work, [illegible] dislike it less than most doctors. Bea[illegible] all administrators spend much of their [illegible] letters and report forms, the work of an o[illegible]

Simi[illegible]gh score on chemist may suggest research in medicine involving chemistry. A high score on the public administrator scale might indicate work in a public medical activity in contrast to private practice. A high score on the life insurance salesman scale, not at all common among doctors, might suggest some sales activity related to medicine.

Edward K. Strong, Jr., Ph.D.
Professor of Applied Psychology

Anthony C. Tucker, Ph.D.
Colonel, Medical Service Corps, U.S. Army

Stanford University, California
January 1952

REPORT OF MEDICAL INTERESTS

NAME ....

ADDRESS .... DATE ....

....

Your interests are similar to those of:

	C No	B− Probably Not	B Possibly	B+ Probably	A Yes
1. Physicians?					
2. Medical specialists?					
And of the four groups of specialists considered: 3. INTERNISTS?					
4. SURGEONS?					
5. PATHOLOGISTS?					
6. PSYCHIATRISTS?					

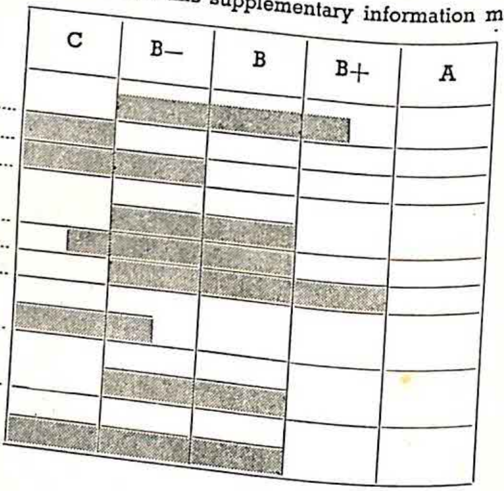
7. Where two or more specialists interests scores are about equal, your interests are more similar to (a) than to (b) as follows:

a. .... b. .... [illegible]

4− Tied	5 to 9 Probably	10+ Yes

8. Your Vocational Interest Blank [illegible] cored on other scales and this supplementary information may be helpful to you.

a. Administrative Interests
    Public Administrator ....
    Office Worker ....
    Personnel Manager ....
b. Teaching and Research Interests
    Math-Science Teacher ....
    Psychologist ....
    Chemist ....
c. Persuasive Interests
    Life Insurance Salesman ....
d. Linguistic Interests
    Author-Journalist ....
e. Aviation Interests
    Aviator ....



9. Comments.

REFERENCES

1. Advisory Board for Medical Specialties. *Directory of Medical Specialists* (Vol. IV). Chicago: A. N. Marquis, 1949.
2. Burnham, P. S. Stability of interests. *Sch. and Soc.*, 1942, 55, 333.
3. Holmen, M. G. Vocational interest patterns of professional specialists. Unpublished Ph.D. dissertation, Stanford Univer., 1952.
4. Kelley, T. L. The scoring of alternative responses with reference to some criterion. *J. educ. Psychol.*, 1934, 25, 504-510.
5. Strong, E. K., Jr. *Vocational Interests of Men and Women.* Stanford, California: Stanford Univer. Press, 1943.
6. Strong, E. K., Jr. Norms for Strong's Vocational Interest Tests. *J. appl. Psychol.*, 1951, 35, 50-56.
7. Strong, E. K., Jr. Permanence of interest scores over 22 years. *J. appl. Psychol.*, 1951, 35, 89-91.
8. Strong, E. K., Jr. Twenty year follow-up of medical interests, Chapter 7 in L. L. Thurstone (Ed.), *Applications of Psychology* (in honor of Walter V. Bingham). New York: Harper, 1952.
9. Tilton, J. W. The measurment of overlapping. *J. educ. Psychol.*, 1937, 28, 656-662.
10. Zuckerman, J. V. Interest item response arrangement as it affects discrimination between professional groups. *J. appl. Psychol.*, 1952, 36, 79-85.

Volume 66
Number 10

Whole No. 342
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Occupational Level Scale as a Measure of Drive

By

GORDON J. BARNETT
New York City
IRVING HANDELSMAN
New York City
LAWRENCE H. STEWART
George Peabody College for Teachers
and
DONALD E. SUPER
Teachers College, Columbia University

Accepted for publication September 11, 1951

Price $1.50

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH ST. N.W., WASHINGTON 6, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PREFACE

What is the meaning of occupational level as measured by interest tests? Is having interests like those of executives a sign of motivation to strive to rise to high position? Or is occupational interest level simply an index of similarity of interests to those of men at various occupational levels, devoid of dynamics and unrelated to drive?

It is these questions which we have sought to answer in the studies and in the final synthesizing essay which constitute this monograph. Barnett and Handelsman were stimulated to test the validity of common interpretations of the Occupational (Interest) Level scale of Strong's Vocational Interest Blank; Stewart was moved, by the lack of such research, to look into the nature of the relationship between adolescent experiences and interests; an[illegible] writer had the pleasure of helping Barnett, Handelsman, and [illegible]e their independently conceived studies in relation to each ot[illegible] organize their results into one systematic report. We are inde[illegible] Drs. Herbert S. Conrad and Edward K. Strong, Jr., for their h[illegible] and suggestions with the manuscript; the interpretations are our own.

The results should be helpful to vocational counselors and personnel workers using tests in counseling and selection. They should be of value to students of developmental psychology and of motivation. They should stimulate further research in an important area of theory and practice.

Teachers College, Columbia University
January, 1951

Donald E. Super

THE OCCUPATIONAL LEVEL SCALE AS A MEASURE OF DRIVE

Preface iii
I. *Introduction* 1
II. *Level of Occupational Interest and Satisfaction with Chronic Vagrancy* 4
III. *The Relationship between Occupational Level (OL) Scores and Various Measures of Aspiration and Achievement* 13
IV. *The Relationship of Certain Social Factors to Occupational Level as Measured by Strong's Vocational Interest Blank* 18
V. *The Meaning of Occupational Interest Level as Measured by Strong's Blank* 22
Appendix A: Unemployment Satisfaction Scale 30
Appendix B: Student's Self-Rating 30
Appendix C: Rating of Level of Aspiration 31
Appendix D: Biographical Questionnaire 31
Bibliography 36

Chapter I

INTRODUCT[illegible]

Level of Occupational Interests: Measure of Status or of Drive?

Vocational interest has often been conceived of by educators, social workers, and psychologists as more or less synonymous with drive, the underlying assumption apparently being that if an individual likes an activity he will seek to engage in it and apply himself to it with as much effectiveness as his ability level permits. When Strong's analysis of the relationship between occupational interests and socioeconomic level (49: Ch. 10) showed that occupational interests can be arranged in a hierarchy according to social level, just as occupations can be classified according to intelligence, it seemed to many that the equivalence of interest and drive had been demonstrated and, furthermore, that Strong's scale for measuring *occupational interest level* provided an instrument for the measurement of drive or aspiration level.

This scale has been the subject of a n[illegible] studies which are discussed bel[illegible] is used by many counselors wor[illegible]h Strong's Vocational Interest Blank. But the few published investigations of the meaning of the Occupational Level scale have not, as will shortly be made clear, shown to what extent level of vocational interest is a driving force which motivates the individual to strive to attain a status compatible with his level of interest, or simply a reflection of socioeconomic level. Is having interests like those of professional men an index of motivation to strive to rise to the higher socioeconomic level? Are adolescents with interests like those of business executives inclined to show, in their school activities and in their educational and vocational goals, patterns of behavior which can be interpreted as indications of a high aspiration level? Are men whose level of vocational interests is high likely to be dissatisfied with a low level of accomplishment, while those with low occupational interest levels are satisfied with unemployment, drifting, and dependency? Or is occupational interest level simply a measure of the extent to which an individual has, through cultural influences in or outside of the home, developed interest in objects and activities which happen to characterize one or another of the various socioeconomic levels? Is it possible, in other words, that there is no drive or motivating force inherent in level of vocational interest? These are important questions to the counselor and personnel worker.

The Occupational Level (OL) Scale

It has been demonstrated that occupations may be arranged in a hierarchy according to the average intelligence of the people engaged in them (47). Prior to the development of the OL [illegible] was not known whether there ex[illegible] lar rank order of occupations [illegible] of interests, and, provided there [illegible]ether the two hierarchies were sim[illegible].

In order to investigate this problem, Strong constructed the OL scale, which contrasts the interests of semiskilled laborers with the interests of business and professional men who were earning approximately $2,500 and up annually during the depression of the 1930's. It was entirely possible that, while the unskilled group differed from the business and professional group in terms of intelligence, there might be men at every occupational level with a given type of interests, because they belonged in the same occupational field, such as semiskilled mechanics, machinists, and professional engineers. Differences between *levels* might reflect ability, whereas differences in interests might affect only *fields*. Instead, Strong found a definite progression in mean OL scores from unskilled to semiskilled to skilled workers, and from there upward to business executives and professional men. He concluded:

> Men cannot be assigned to a specific occupation on the basis of intelligence test scores. Neither can they be assigned to a specific occupation in terms of an OL score. In both cases there is far too great overlapping between occupations to make this possible. Nevertheless, such assignments can be made in a general way; men with high OL scores have the interests of business executives and professional men, but those with low scores have the interests of workmen (49, p. 195).

After extensive counseling experience with Strong's Blank, Darley drew the following conclusions concerning the meaning of the OL scale:

> Occupational level, a quantitative statement of the eventual adult "level of aspiration," represents the degree to which the individual's total background has prepared him to seek the prestige and discharge the social responsibilities growing out of high income, professional status, recognition, or leadership in the community; at the lower end of the scale the individual's background has prepared him for the anonymity, the mundane round of activities, and the "followership" status of a great majority of the population (17, p. 60).

Elsewhere Darley has concluded:

> I am becoming clinically more convinced . . . that the occupational level score is an excellent quantitative statement of a form of motivation or level of aspiration or some other dynamic aspect that is . . . referred to in the literature as "drive" (49, p. 465).

Studies of the Meaning of the OL Scale

There have been few attempts to investigate the factor or factors measured by the OL scale. Most of the present knowledge is based on Strong's data collected in the development of the scale. Other studies that have been made show little agreement with respect to findings. These are referred to when relevant in the final chapter of this monograph.

What Is the Significance of Occupational Interest Level?

The three chapters immediately following describe three studies, each of them designed to investigate the nature of the factors measured by the OL scale. While all three studies were carried out under the sponsorship of one person, each represents an independent and distinctive approach to validation of the OL scale. While there is enough agreement among the findings to make possible some definite conclusions, there are some discrepancies which still challenge the counselor and researcher. The final chapter considers the three studies, synthesizing the compatible conclusions of these and earlier investigations and attempting to ascertain logical grounds for the dissimilarities.

CHAPTER II

LEVEL OF OCCUPATIONAL INTEREST AND SATISFACTION WITH CHRONIC VAGRANCY*

GORDON J. BARNETT

Problem and Method

If it were discovered that, of a group of subjects who had no jobs and had not had any job for much of their lives, those who were satisfied with being unemployed earned low OL (occupational level) scores and those who were dissatisfied earned high OL scores, and if it were further found that there existed a high negative relationship between intensity of satisfaction and OL scores, then it might be reasoned that OL scores do provide indices of people's levels of vocational aspiration. This proced[illegible]s employed, in a study of chronic [illegible] in an attempt to learn more a[illegible] is really measured by the OL s[illegible]

Source of Data: Chronic Vagrants

From a group of 37,644 applicants for food and/or shelter at the Salvation Army Men's Social Service Center, located in the Hell's Kitchen district of New York City, 763 men were interviewed for biographical data and detailed occupational histories. Selection was made at the rate of approximately 1 applicant in 50, the refusal of 43 men to cooperate reducing the rate slightly below that figure. All veterans, the obviously physically and mentally incapacitated, and, in one subgroup studied, all men younger than 30, were excluded in order to eliminate those whose chronic unemployment might be considered justified and to derive a sample of real vagrants.

All the applicants who were selected for study were designated chronically or nonchronically unemployed, according to the following criterion. The total possible work life of each applicant was determined by means of detailed questioning. Total possible work life was considered to be the total time in months that had elapsed since the applicant had finished his formal education and left his home (or substitute therefor). An index of unemployment was computed for each applicant interviewed by dividing his total possible work life in months into total months unemployed. If an applicant had left school at the age of 14, his work life was considered to have begun at that time; if he had earned a master's degree, and had been steadily engaged in study up to the date the degree was granted, his work life was considered to have begun with the awarding of the degree. Work done during summer vacations while the individual was still technically a student was disregarded, regardless of whether he had then completed the sixth grade or had earned a graduate degree.

When the applicant's index of unemployment had been determined, he was classified as chronically or nonchronically unemployed according to whether his index was more or less than 51, i.e., whether he had been unemployed for more or less than half his possible work life, the former, of course, being the chronics. All chronics were instructed to move to a small room adjoining the room in which they had been interviewed, where, under conditions more suitable for the administration of psychological tests, they were administered the measuring devices described after the personal history data in the following section.

---

* Part of a Ph.D. investigation conducted at Teachers College, Columbia University, under the direction of Dr. Donald E. Super, Dr. Laurance Shaffer, Dr. Irving Lorge, and Dr. Robert L. Thorndike.

Factors Investigated

As each man was interviewed, this information was elicited:

Age in months. A person's age provides an indication of his maturity and of the degree to which he has been exposed to his culture. Physiological and psychological changes occur with age which might be expected to modify one's attitudes. In the case of the men under investigation in this study, age might conceivably have either of two diametrically opposite effects on attitude toward chronic unemployment. First, older men might be discovered to be generally more satisfied than the younger men as the result of having become more or less inured to unemployment and the hardships inherent in it, with the younger men not as yet accustomed to the rigors of such a life and thus generally less satisfied with it. Second, it might be found that the younger men are generally more satisfied, as a result of not yet having come to grips with the multitude of problems that face anyone who is unwilling to work in our society, with the older men more disillusioned, more bitter, about such an unproductive and unrewarding existence, and hence generally less satisfied with it. While the stereotype is one of the conservative old man and the radical young man, still it is not inconceivable that this latter relationship might be found to exist. Therefore, as closely as was possible with a group of this type, each applicant's age was determined.

Educational level. While there is great variation in the amount of learning by individuals during a given period of schooling, the level of education attained does provide some evidence of one's accomplishments. The relationship that seems most logical appears to be one in which the better educated are more dissatisfied with the life of a vagrant. Experience at the Center had indicated no such general relationship, however, the better educated applicants frequently expressing satisfaction with their way of life, while poorly educated applicants expressed dissatisfaction at times. In order to check more thoroughly on this factor, education in months, represented by the highest grade reached, was ascertained.

Father's occupation. Occupation is one index of the degree of success attained in the working world. In this country, at least, occupation represents a widely accepted criterion of success. The occupations of the applicants' fathers (major occupations in the cases of fathers who had held more than one type of job) were used as an index of parental socioeconomic status in this study. It could be expected that applicants from "better homes" in terms of their fathers' occupations would be less satisfied with the life of the vagrant than those from homes in which the nominal head of the family had never risen above the level of the majority of people about him. In view of the frequent discrepant relationships discussed above, however, father's occupation was included for study.

Following the direct questioning, each man who met the above-mentioned criteria for designation as a "chronic" vagrant was administered the written test materials, which covered these areas:

Intelligence. Scores on verbal intelligence tests provide reasonably valid indicators of capacity for abstract thought. Level of intelligence may [illegible]cted to influence attitudes; in the case [illegible]de toward being chronically une[illegible]ould seem that the most logical rela[illegible]ould be one in which the more inte[illegible]n are least satisfied, while the less [illegible]t are most satisfied. However, experience with applicants at the Social Service Center had demonstrated that the expected relationships did not hold here any more consistently than in the case of age, education, or father's occupation. Frequently men of obviously modest intelligence were loudest in their denunciation of their way of life, while an occasional college graduate, perhaps a vagrant since shortly after his graduation from college, might profess to love his way of life and to harbor not the slightest desire to change it. This impressionistic evidence of a discrepant relationship between intelligence and satisfaction with unemployment had the disadvantage of being based solely on observation of the men's behavior and on their orally expressed statements of satisfaction and dissatisfaction. Consequently this study investigated intellectual functioning by means of a standardized intelligence test and satisfaction by means of the Satisfaction Scale described below.

The intelligence test used was the Otis Self-Administering Test of Mental Ability, Higher Form D. This test consists of 75 items, requiring a certain amount of verbal facility and ability to see logical relationships.

The Higher Form of the Otis is intended for high-school students and college freshmen.

It was originally planned to use the Intermediate Form, designed for grades 4 to 9, in this study. However, in a previous study the educational levels of applicants at the Center suggested using the Higher Form, the mean grade completed being slightly above the tenth grade.

Interests. The Strong Blank was completed according to the instructions printed on the blank. The time required for completion of the 400 items ran from 15 minutes to about one hour, with a mean of approximately 25 minutes. A pilot study attempted to determine whether difficulty in comprehension would be presented by certain items. No such difficulties were encountered in 20 cases so studied, the men apparently understanding names and terms used as well as might be expected of a group of this educational level.

Satisfaction. The most important single factor subjected to investigation in this study was the relative satisfaction or dissatisfaction of the applicant with his vagrancy. The most promising method of determining this relative satisfaction or dissatisfaction appeared to be by means of a series of scaled responses, ranging from those indicative of utter dissatisfaction to those signifying complete satisfact [illegible] such scale existed, and while it wou [illegible] een possible to construct one, the l [illegible] d appeared to be disproportionate [illegible] antages such a scale might offer.

A more practical approach sugg [illegible] lf in the form of Hoppock's Job Satisfaction Blank (31). This blank has been widely used and has been proved to be of value in the determination of job satisfaction or dissatisfaction. It consists of four scales, made up of from six to ten items, the responses to which are totaled in order to obtain a composite index of satisfaction. In this manner it is possible to express the degree of satisfaction or dissatisfaction in quantitative terms.

This blank was reworded to be pertinent to "having-no-job" satisfaction instead of to job satisfaction, retaining the original format and altering the wording only to the extent required to make the items germane to unemployment (Appendix A). Use of a technique such as this implies two assumptions, first, that the items within any one scale are equidistant, and second, that the four scales are equally important and equally valid. These conditions are but partially met in any revision. This is why extreme groups of scores on the revised satisfaction blank were selected for comparative study. In order to allow for a tendency to mark the first responses that expressed approximately how the applicant felt, the responses indicating satisfaction were placed first in two of the scales and last in the other two.

The Nonchronic Test Group

For each applicant who was found to be a "chronic" and was tested, a "nonchronic" was administered the same battery of tests. The first nonchronic who appeared each time a chronic was located among the incoming population was selected for testing. A total of 175 nonchronics completed the battery, against a total of 174 chronics. For reasons not satisfactorily explainable, since the task involved in the intelligence test was presented to them in the same manner, none of the nonchronics was unwilling to complete the Otis, whereas 23 chronics refused to do so.

Treatment of the Data

Otis tests were hand scored according to the instructions in the test manual. Total score, the number of correct answers, was checked twice in order to insure that no computational errors had been made.

The Strong Blanks were hand scored, rescored by a second person, and in the case of a discrepancy between first and second scoring were scored a third time. The final score represented a compromise between the third score and the one closest to it when there was not perfect agreement with one of the other two. All the Strong data were treated in raw score form.

Responses to the Satisfaction Blanks were tallied and a composite "index of satisfaction" was obtained. The item responses indicating the greatest degree of dissatisfaction were assigned a weight of 1. The others were given values in order, from 2 to *n*, the latter representing the total possible number of items in a particular scale. It was then possible to obtain a scale value in terms of the standard deviation. Deviations in the direction of dissatisfaction were given a minus weight. Those in the direction of satisfaction were assigned a plus weight. In this manner it was possible to reduce the scores on each of the four scales to comparable units, and the scale values of the items checked were added to obtain the index of satisfaction.

Fathers' occupations were quantified by means of the Barr Occupational Scale. This scale was originally employed by Barr and Terman (3) in studies of gifted children, and has been widely used since 1925. It assigns numerical scores to well-known occupations, leaving adequate space between the scores for the interpolation of other occupations not mentioned in the scale. In the case of occupations not specifically mentioned in the Barr Scale, a panel of five psychologists, all with special knowledge of occupational requirements, arrived at a Barr score for each occupation, based upon estimates of the level of intelligence required for placement in the occupation. For the not-so-well-known occupations, reference was made to printed job descriptions, to occupational monographs, and to the Dictionary of Occupational Titles (57).

Personal Data

Chronics. Barr Scale scores of the fathers of the chronic group ranged from 3.62 (laborer) to 18.10 (president, insurance company). The mean score fell approximately one-half the distance between the occupations of potter and tailor. Seven of the applicants were unable to specify what their fathers' occupations were. None of these seven had any idea who his father was.

The mean age of the men in the chronic group was 46.97 years. Most studies have found the mean age of vagrants to be in the vicinity of 20 years, with a steady decrease in the proportion of older men. The four applicants who were under 30 were excluded from this group. The total range of ages was from 31.33 to 66.50 years.

Most surprising was the high average educational level claimed. Previous investigators have noted a tendency for vagrants to exaggerate their past accomplishments, including their education. A comparison of the statements of these vagrants with those made in other studies reveals, however, that whereas the latter groups usually averaged five or six years of school, the chronics in this study claimed a mean level approximating the tenth grade. There is no obvious reason why these men would have exaggerated their educational accomplishments any more than previous groups studied. It therefore seems reasonable to infer that these subjects had a higher average level of education than other groups studied. This may well be in part a reflection of the general increase in education (two grades) in the adult population in this country during the interim between World Wars I and II. The range of claimed education in the chronic group was from the fourth grade to the Ph.D. degree. The latter claim was investigated and verified.

Indices of unemployment ranged from 51 to 95, indicative of unemployment from one-half to more than nine-tenths of total working life. (The [illegible]on for designation "chronic" was a his[illegible]ving been unemployed during more [illegible]f of one's total possible working li[illegible] applicant earned an index of 100, it [illegible]t that even the "most-chronic" of the [illegible] claimed some employment. Only two [illegible]cs earned an unemployment index greater than 86.

Nonchronics. Fathers' occupations in the nonchronic group ranged from laborer to lawyer. As in the case of some chronics, some (25) of the nonchronics were unable to specify the occupations of their fathers for similar reasons. Mean Barr score of the fathers of the nonchronics was 8.05, equivalent to the occupation of carpenter.

Because men younger than 30 were not excluded from the nonchronic groups, the range of age was greater, from 21.25 to 67.18 years. Mean age was 45. The upper extremes of the chronic and nonchronic groups were approximately equal, and in the two groups the percentage of men older than 50 was about the same, 32 and 37 per cent, respectively.

Again, as in the chronic group, the men claimed a surprisingly high educational level. The average nonchronic claimed to have completed approximately ten years of school. The total range of grades was somewhat greater in this group, from no formal education (claimed by four men) to the M.A. degree (one man).

Indices of unemployment in the nonchronic group ranged from 1 to 50, with a mean of

33.94. This is indicative of a history of almost no unemployment during total possible working life to a history of having been unemployed during exactly one-half of one's working life, as nearly as could be determined. Eighty-three per cent of this group earned indices between 25 and 50, indicating that even among the nonchronics there were few who had histories of reasonably steady employment.

Test Data

Chronics. The chronics' mean Otis raw score was 44.44, the equivalent of an IQ of 102, a percentile rank of 55, and a classification of "normal" according to Otis norms. Raw scores ranged from 23 to 72, representative of IQ's from 81 to 130 and of classifications from "dull" to "superior." These intelligence test results served at least in part as a check on the veracity of the statements of the applicants concerning their educational and socioeconomic backgrounds. It would appear highly doubtful that the mean score of 44 could be earned by any group without its constituents' possessing the intellectual ability that such a score implies. When it is borne in mind that the Otis was designed for and standardized upon [illegible] school students and college freshmen, it [illegible] apparent that the applicants not [illegible]d surprisingly high on the test, bu[illegible] exceeded the mean scores of a gro[illegible]g a distinct advantage with respect to [illegible]

Indices of satisfaction ranged from —6.48 to 8.93, the former indicating that all four extreme statements in the direction of dissatisfaction were checked, the latter indicating that all the extreme statements in the direction of satisfaction were marked. The mean score of .063 indicates that so far as the Blank can disclose, the chronic group tended to be neither markedly satisfied nor strongly dissatisfied, although 8 men of the 174 indicated complete satisfaction as measured by the Blank.

OL scores ranged from —129 to 206, with a mean raw score of —.158, and a standard deviation of 55.02. Strong's one thousand men ranged from —200 to 240, with a mean raw score of 18.1 and a standard deviation of 73.62 (49, p. 189). Thus, the chronics averaged about 18 raw score points below Strong's criterion group. Strong (49, p. 291) reports that 62 unemployed men in the Palo Alto shelter earned a mean OL score of approximately zero, comparable to the mean score of the chronic group.

Nonchronics. The nonchronics earned a mean raw score of 46.44 on the Otis. The scores ranged from 17 to 73, a range slightly greater than that observed in the chronic group. Their mean score is equivalent to an IQ of 104, a percentile rank of 63, and a classification of "normal." The standard deviation of the Otis raw scores is 10.39.

Owing to a misinterpretation of Otis' instructions by one of the two persons responsible for the administration of the Otis tests, the chronics were not limited to a 30-minute time limit as the nonchronics were. The estimated mean time required by the former was between 25 and 40 minutes. Consequently, direct comparison of intelligence test scores of the two groups is misleading. Nevertheless, the close equivalency in mean scores and in range and standard deviation suggests that under more closely equated conditions little difference would be found between them.

Indices of satisfaction ranged from —6.48 to 8.93, there having been one man who expressed complete satisfaction and ten who indicated complete dissatisfaction. The mean index of —.554 indicates, as in the case of the chronic group, that the average nonchronic was neither strongly satisfied nor dissatisfied with his vagrancy, tending to be relatively indifferent in terms of the Satisfaction Blank.

The nonchronics showed a slightly narrower range of OL scores, the extremes being —135 and 181. The mean raw score was 28.44, approximately equivalent to the occupation of printer, and fell about ten raw score points above the mean of Strong's criterion group.

Results

Table 1 and 2 present the intercorrelations of all factors in the chronic and nonchronic groups, respectively.

Chronics and Nonchronics

Chronics. It will be recalled that a negative relationship between OL and degree of satisfaction with vagrancy was hypothesized, providing the OL score is related to motivational and/or aspirational factors. This hypothesis was supported in the chronic group by our finding a product-moment correlation of —.73 between OL and index of satisfaction. In order to investigate the influence of the other factors included for study, i.e., education, intelligence, age, father's occupation, and index of unemployment, first-order correlations were computed for OL and index of satisfaction, holding the other factors constant. These partial cor-

TABLE I

INTERCORRELATIONS OF ALL FACTORS: CHRONICS

	N	2	3	4	5	6	7
(1) Occupational Level	174	−.73*	.14	−.06	.19†	−.10	.28*
(2) Index of Satisfaction	174		−.19*	.05	−.26*	.12	−.23*
(3) Father's Occupation	167			−.12	.30*	−.14	.59*
(4) Age	174				.10	−.09	−.22*
(5) Education	174					−.14	.49*
(6) Index of Unemployment	174						−.06
(7) Intelligence	151						

* Significant at 1% level.
† Significant at 5% level.

relations indicated that the other factors exerted little influence on the zero order correlation of −.73, the greatest divergence being observed in the case of intelligence, which, when held constant, resulted in a first-order correlation of −.78.

OL and age were found to correlate −.06, a figure that is statistically insignificant. Strong (49, p. 189) presents conflicting information concerning his findings with respect to OL and age. For his one thousand men in general, Strong found a slight increase in OL with increasing age, but no change in follow-up data on high-school juniors. He concluded that an increase of from two to three OL standard score points might occur with increasing age, but that age affects OL very little.

Claimed education and OL correlated .19 in the chronic group and −.19 in the nonchronic group. These correlations are significant at the .05 level. The reason for the discrepancy is not apparent. Strong does not present data concerning the relationship between education and OL. Indirectly he makes reference to it in his discussion of the relationship between age and OL, for in his younger groups there was, of course, a direct relations[illegible] between age and grade in school.

[illegible]ationships between index of t[illegible] and other factors were gener[illegible]xpected. Correlations of −.19, −.[illegible] −.23, all significant at the .01 level, were found between satisfaction on the one hand and father's occupation, claimed education, and intelligence on the other. Thus, chronics claiming higher fathers' occupations tended to express greater dissatisfaction with unemployment. Similarly, the more intelligent men and those claiming higher educational attainments were less satisfied with unem-

TABLE 2

INTERCORRELATIONS OF ALL FACTORS: NONCHRONICS

	N	2	3	4	5	6	7
(1) Occupational Level	175	−.68*	−.24*	−.30	−.19†	−.39*	−.24*
(2) Index of Satisfaction	175		−.41*	.07	−.20*	.46*	−.17†
(3) Father's Occupation	564			−.09†	.26*	−.10*	.65*
(4) Age	589				−.14*	−.07	−.31*
(5) Education	589					.10	.44
(6) Index of Unemployment	589						−.02
(7) Intelligence	175						

* Significant at 1% level.
† Significant at 5% level.

ployment than the less intelligent and not-so-well educated.

Strong (49, p. 201) reports a correlation of .05 between OL and intelligence as measured by the Thorndike Intelligence Examination for High School Graduates. The correlation of .28 between OL and Otis scores in the chronic group indicates a closer positive relationship. Claims concerning educational accomplishments and fathers' occupations received some support from correlations of .59 and .49 between these two factors and intelligence in the chronic groups, and of .44 and .65 in the nonchronic groups.

Nonchronics. The correlation between OL and index of satisfaction in the nonchronic group was −.68. Partial correlations were computed, as was done in the case of the chronics. Again, the other factors were shown to be contri[illegible] little to the observed correlation [illegible] est divergence being noticed [illegible] ex of unemployment was held c[illegible], resulting in a correlation of −.6[illegible]ween OL and satisfaction.

OL and age correlated −.30 in the nonchronic group, significant at the .01 level. In view of Strong's conclusion that age has little effect on OL and our having found a correlation of −.06, not signigficant at the .05 level, in the chronic group, this figure poses a problem concerning the relationship between OL and age. What changes Strong did note in connection with OL and increasing age were in an upward direction, while our negative correlation would indicate that the older nonchronics tended to earn *lower* OL scores.

The implication would seem to be that at least some of the nonchronics were individuals who were still trying to obtain employment, but as they grew older the level of their ambitions fell lower and lower. The chronics, on the other hand, were largely individuals who had renounced any claim to, or interest in, employment. Thus, the passing years had not resulted in their lowering their vocational ambitions. Strong's finding of a slight increase in OL with advancing age was based on average people who were presumably fairly well adjusted to their work; thus their aspirations tended to go up rather than down.

A correlation of −.30 was found between OL and father's occupation in the nonchronic group. This correlation is significant at the .01 level, contrasted with the statistically insignificant correlation of .14 in the chronic group. The negative correlation suggests a partial answer to the question of why these men became vagrants. The men claiming higher fathers' occupations should be expected to earn higher OL scores, there being a positive relationship between OL and intelligence, and between father's occupation and intelligence. The negative correlation suggests that while the men from the relatively better homes suggested by the higher fathers' occupations did tend to be more intelligent than men claiming lower fathers' occupations, they were, nevertheless, handicapped to a certain extent by having interests not commensurate with their socioeconomic levels. This may explain their having failed to become useful members of society, the factor or factors implied by a low OL score having operated in their becoming vocationally and socially maladjusted.

Satisfied Versus Dissatisfied

In order to point up the relationships among the factors studied, the chronic and the nonchronic groups were both divided into "Satisfied" and "Dissatisfied" subgroups. This was done by designating those whose satisfaction indices fell above the third quartile of each distribution as "satisfied" and those whose indices fell below the first quartile as "dissatisfied." The score limits in the chronic group were 2.00 and above for the satisfied and −2.42 for the dissatisfied. Similar critical scores in the nonchronic group were 1.70 and −2.91.

The critical ratio was employed to test the significance of the differences between various scores of the chronics and nonchronics, as well as between scores of the satisfied and dissatisfied subgroups. Two factors differentiated the chronics from the nonchronics at the .01 level. The latter earned significantly higher OL scores and had higher fathers' occupations. None of the other factors, i.e., age, claimed education, intelligence, or index of satisfaction, differentiated between the chronics and the nonchronics.

The same two factors, OL and father's occupation, showed differences significant at the .01 level between the satisfied and dissatisfied nonchronics. The satisfied and dissatisfied chronics differed significantly with respect to OL and claimed education. Using the quartile extremes, as described above, did not demonstrate significant differences in claimed fathers' occupations in the chronic group.

Applying more stringent criteria for designation as satisfied and dissatisfied, all the distributions of satisfaction indices were recalculated, using the upper and lower deciles instead of quartiles. With this new criterion in effect one change was observed: fathers' occupations claimed by the satisfied and dissatisfied chronics now differed at the .01 level of confidence.

Thus two factors, *father's occupation* and *OL score,* differentiated not only between the chronics and nonchronics but also between satisfied and dissatisfied members of both groups. In only one instance, that of the satisfied and dissatisfied chronics, did another factor (claimed education) differentiate significantly between these paired groups. None of the other factors approached the level of significance observed in the case of OL and father's occupation.

While these two factors differentiated significantly between the dichotomies established, the sizes of the correlations between them in the chronic and nonchronic groups, .14 and −.24 respectively, ruled out any assumption that one could predict the occupation of a vagrant's father from a knowledge of his OL score. It follows that the factors represented by fathers' occupations, particularly socioeconomic background, and by the OL scal[illegible]ted separate, albeit similar, in[illegible] the men studied. Not only c[illegible] chronic vagrants be distingu[illegible]om the nonchronics on the basis [illegible] father's occupation and OL score, but within these two groups the satisfied and the dissatisfied could also be isolated.

Conclusions

Negative correlations of −.73 and −.68 were found between OL score and index of satisfaction with unemployment in the chronic and nonchronic groups respectively. This might lead one to accept the hypothesis that the OL score provides a measure of level of aspiration, at least vocationally speaking. As OL scores increased in magnitude, expressed satisfaction with vagrancy declined.

However, another factor, father's occupation, also was negatively correlated with satisfaction with unemployment and differentiated between the satisfied and dissatisfied men in both the chronic

(−.19) and nonchronic groups (−.41) as well as between the groups themselves. The subjects' expressed dissatisfaction with unemployment should not, therefore, be viewed in terms of high OL scores alone. Nor could the nonchronics be distinguished from the chronics solely on the basis of their OL scores. In each case, father's occupation also entered the relationship. Furthermore, OL scores and fathers' occupations were not significantly correlated, thus ruling out the possibility of equivalency of OL score and father's occupation. In fact, in the nonchronic group OL scores and fathers' occupations were significantly *negatively* correlated. This suggests that possession of a level of occupational interest not commensurate with original socioeconomic level may be responsible in part for social and vocational malad[illegible]ent.

These data cannot be said to have demonstrated that the OL scale measures drive. They have shown that it is related to status satisfaction, and that it is not simply a measure of parental socioeconomic level. While it apparently operates in the same way as the latter, it is independent of it. Perhaps the OL scale measures just what its method of construction suggests: the occupational level to which the individual's interests correspond. The scale might therefore be of use in a guidance or selection battery as an index of the occupational and/or social level which the individual is likely to find congenial. It is, in a sense, a measure of satisfaction or dissatisfaction with a way of life. If it is also a measure of drive, this remains to be demonstrated.

THE RELATIONSHIP BETWEEN OCCUPATIONAL LEVEL (OL) SCORES AND VARIOUS MEASURES OF ASPIRATION AND ACHIEVEMENT*

Irving Handelsman

In the present experiment, the OL score (Strong's Vocational Interest Blank), laboratory tests of level of aspiration, and teachers' ratings and students' self-ratings of level of aspiration are used as measures of level of aspiration. The extent of student participation in extracurricular activity and three-year high-school grade averages are used as measures of achievement which should, on a priori grounds, be affected by aspirational level.

The OL scores may be hypothesized as reflecting the individual's socioeconomic goals in life. The teachers' ratings and the students' self-ratings of level of aspiration are considered to be indices of the manifest motivation and striving of the student. The extent of student participation in extracurricular activity and high-school grades are believed to depend, in part, upon the student's general level of aspiration in his daily living, and are behavorial in nature. On the other hand, it is believed that the laboratory tests of level of aspiration are more specific and less related to life situations, and measure the individual's aspirational level in tasks of limited significance. There is a dichotomy in these measures of level of aspiration, with one group representing life situations in the broad sense whereas the other group is unrealistic and narrow in scope.

Problem and Method

The present study proposed to determine to what extent the above-mentioned, presumed measures of level of aspiration are related to each other as well as to other selected factors. The aspect of the problem reported here is to determine the degree of relationship between the OL score of the Strong Vocational Interest Blank and various other presumed measures of level of aspiration, such as teachers' ratings of students' level of aspiration, students' self-ratings of lev[illegible]spiration, and two laboratory t[illegible]vel of aspiration. Secondly, an [illegible] made to ascertain whether any re[illegible]p exists between the OL scor[illegible]d scholastic and extracurricular achievement. The third part of the study, not reported here, was concerned with an investigation of the nature and meaning of two laboratory tests of the level of aspiration.

Design of the Study

Sample. The experimental group consisted of 132 white male students who were members of the junior and senior classes of two high schools in a suburban community located about 30 miles from New York City. The schools are referred to as "School A" and "School B." Sixty-four subjects participated in this study from School A, while 68 students participated from School B. The average age of the students from School A is seventeen years and nine months (*SD* 9.0 months) while the average age of the students from School B is seventeen years and ten months (*SD* 8.8. months). The average IQ, derived from the Otis Self-Administering Test of Mental Ability, Higher Examination, Form A, is 105.7 (*SD* 10.8) in School A and 107.6 (*SD* 10.1) in School B.

---

* Part of an Ed.D. investigation conducted at Teachers College, Columbia University, under the direction of Dr. Donald E. Super, Dr. Laurance F. Shaffer, Dr. Nicholas Hobbs, and Dr. Irving Lorge.

The laboratory tests of level of aspiration. The two laboratory tests used for measuring the level of aspiration were designed by the writer. The verbal test consisted of paragraphs in which the subjects were asked to read the material and to put a line through all the written words which were numbers. The arithmetical tasks involved the computation of simple three-digit, three-row problems in addition. Copies of the instructions and the tests are contained in the original manuscript (27).

The range of intercorrelations of the five performance trials of the [illegible]etic level of aspiration test is from [illegible] with a median coefficient of cor[illegible] .68 in School A, while the ran[illegible]ercorrelations is from .76 to .88 w[illegible]a median coefficient of correlation of .84 in School B.

The rating scales of level of aspiration. Two five-point rating scales of level of aspiration were designed (Appendices B and C). The scales were identical, except that the teacher rating scale was descriptive of the students' drives and was written in the third person whereas the students' self-rating scale was written in the first person. The descriptions ranged from ". . . sets no goals . . ." to ". . . sets high goals. . . ."

Procedure

The intelligence quotient. The IQ equivalent was obtained as recommended by the manual from the Otis Self-Administering Test of Mental Ability, Higher Examination, Form A, which is administered routinely by the respective schools to their students.

The OL score. The Strong Vocational Interest Blank was administered to the subjects by the experimenter and scored by machine.

The level of aspiration scores. The level of aspiration was obtained for each test by finding the mean differences between the performance scores and aspiration scores for the five trials. The average level of aspiration score was computed by taking the mean of the verbal and arithmetical aspiration scores. Since the students were asked to count the number of problems they had completed and record them, a check was made to ascertain the degree of accuracy with which the students scored themselves. A check of the ten trials for the 132 students revealed that 6 erred, once each, in recording the proper score. Three students underestimated their scores by one, and 3 overestimated their scores by one. It appears that these errors were unintentional and that the differences are probably due to chance errors.

Instructions for student and teacher ratings. In each school the teachers who were assigned to rate the students participating in the study met with the experimenter and were instructed to "check the item which most appropriately describes the level of aspiration or drive of the student in his everyday life." A discussion followed the instructions, and it appeared that there was a common understanding as to the meaning of the terms.

The students participating in the experiment were instructed to "check the item which you feel most appropriately describes you in your everyday life." A few questions were raised by the students, but there seemed to be a general understanding of the terms.

Academic achievement. Academic achievement was determined from the school grade averages. Six semesters of work were used for the junior classes, and seven semesters in the case of the seniors. The percentage method of grading students is employed in both schools.

The Extracurricular Activity Index. The Extracurricular Activity Index was obtained by asking the students to fill out a form on which they recorded the names of any school organization, club, or team to which they belonged. In addition, they were told to indicate the number of terms they were members and the number of terms they held office. One point was given for membership and an additional point for each term that an office was held. The sum total of points yielded the Extracurricular Activity Index.

Results

Since the study was conducted in two different schools, the null hypothesis is set up in the case of each variable for each school. It is assumed that there is no real difference between the means of

TABLE 3

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS FOR EACH OF THE VARIABLES IN SCHOOL A ($N=64$)

Variable	1	2	3	4	5	6	7	8	9	10	Mean	Sigma
1. OL Score	—	.42*	.44*	−.004	.42*	.26†	.05	.17	.05	.25†	48.6	6.9
2. Three Year H.S. Average	.42†	—	.41*	.39*	.63*	.33*	.19	.20	.21	.65*	78.4	7.2
3. Student's Self-Rating of Level of Aspiration	.44†	.41*	—	.07	.41*	.32*	.14	.28†	.18	.33*	3.48	.85
4. Age‡	−.004	.39*	.07	—	.28†	−.02	−.05	−.04	.01	−.19	212.7	9.0
5. IQ	.42*	.63*	.41*	.28†	—	.11	−.001	.25†	.08	.46*	105.7	10.8
6. Extracurricular Activity Index	.26†	.33*	.32*	−.02	.11	—	.03	−.16	−.05	.35*	26.9	26.7
7. Verbal Level of Aspiration	.05	.19	.14	.05	−.001	.03	—	.35*	.94*	.30†	64.6	26.0
8. Arithmetic Level of Aspiration	.17	.20	.28†	−.04	.25†	−.16	.35*	—	.62*	.23	57.1	9.4
9. Average Level of Aspiration	.05	.21	.18	.01	.08	.05	.94*	.62*	—	.29*	61.6	14.8
10. Total Teachers' Ratings of Students Level of Aspiration	.25†	.65*	.33*	−.19	.46*	.35*	.30†	.23	.29†	—	8.7	3.4

* Significant at the .01 level of confidence.
† Significant at the .05 level of confidence.
‡ Age measured to nearest month.

the two populations, that any differences are nonsignificant and might easily arise from sampling fluctuations. The critical ratio is computed for each variable and the level of confidence selected for rejecting the null hypothesis is .01.

The two school groups are evenly matched as to OL scores, academic achievement (high-school average for three years of work), students' self-rating of level of aspiration, age, intelligence, discrepancy scores on the arithmetic and verbal laboratory tests of level of aspiration, and total teachers' ratings of students' level of aspiration. There is a significant difference in the extent of partici[illegible] in extracurricular activity on [illegible] of students in the two schools. [illegible] presented in Tables 3 and 4.

O[illegible]re as Related to the Various Measures of Level of Aspiration

OL scores and teachers' ratings. The mean OL score for School A is 48.5 (*SD* 6.9) and for School B it is 48.8 (*SD* 6.8). The correlation coefficient between the

TABLE 4

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS FOR EACH OF THE VARIABLES IN SCHOOL B ($N=68$)

Variable	1	2	3	4	5	6	7	8	9	10	Mean	Sigma
1. OL Score	—	.30†	.04	−.15	.20	.03	.18	.06	.16	.24†	48.8	6.8
2. Three Year H.S. Average	.30†	—	.41*	−.21	.50*	.20	.11	−.02	.08	.71*	77.2	6.8
3. Student's Self-Rating of Level of Aspiration	.04	.41*	—	−.24†	.30†	.38*	−.04	−.05	−.05	.36*	3.2	1.1
4. Age‡	−.15	−.21	−.24†	—	−.46*	.02	.07	−.03	.04	.02	213.6	8.8
5. IQ	.20	.50*	.30†	−.46*	—	.22	.31*	.19	.29†	.29†	107.6	10.1
6. Extracurricular Activity Index	.03	.20	.38*	.02	.22	—	−.15	−.16	−.17	.17	17.8	17.01
7. Verbal Level of Aspiration	.18	.11	−.04	.07	.31*	−.15	—	.66†	.97*	.12	72.5	19.4
8. Arithmetic Level of Aspiration	.06	−.02	−.05	−.03	.19	−.16	.66*	—	.82	.04	58.9	8.4
9. Average Level of Aspiration	.16	.08	−.05	.04	.29†	−.17	.97*	.82*	—	.11	65.8	12.9
10. Total Teachers' Ratings of Students Level of Aspiration	.23†	.71*	.36*	.02	.29†	.17	.12	.04	.11	—	10.3	3.0

* Significant at the .01 level of confidence.
† Significant at the .05 level of confidence.
‡ Age measured to nearest month.

OL scores and the teachers' ratings of their students' levels of aspiration is .25 for School A and .23 for School B, both significant at the .05 level of confidence.

OL score and students' self-ratings. Considerable difference is noted in the correlation of scores with students' self-ratings of aspirational level. In School A the correlation coefficient is .44 ($P = .01$), and in School B it is .04.

OL score and the experimental tests of level of aspiration. The relationships between OL scores and the laboratory tests of level of aspiration are not significant. The correlation between the verbal level of aspiration tests and OL scores is .05 in School A and .18 in School B. The correlation between the arithmetic level of aspiration test and OL scores is .17 in School A and .06 in School B.

Discussion. The OL scores a[illegible]ositively related to teachers' ratin[illegible]dents' level of aspiration in both[illegible]s, whereas they are significantly rel[illegible] to students' self-ratings of aspirational level in School A but not in School B. It appears that there is considerable difference in objectivity of students' concepts of their level of aspiration in both of these schools. OL scores are not significantly related to laboratory tests of level of aspiration.

Score as Related to Academic Achievement, Intelligence, and Extracurricular Activity

OL score and IQ. The correlation between OL scores and IQ is .42 ($P = .01$) in School A, and .20 ($P > .05$) in School B.

OL score and academic achievement. Another significant relationship is that between OL scores and academic achievement. The coefficient of correlation between these two factors is .42 ($P = .01$) in School A and .30 ($P = .05$) in School B. When intelligence is held constant, the correlation becomes .24 and .22 respectively. It is found that the multiple correlation between academic achievement and OL scores and IQ is .66 in School A and .54 in School B.

OL score and Extracurricular Activity Index. The correlation between participation in extracurricular activities and OL scores is .26 ($P = .05$) in School A and .03 in School B.

Discussion. The above findings are in agreement with those of Wren (58), Bradley (8), and Stubbins (50), who, using different research techniques, found intelligence and academic achievement to be highly related to an individual's level of occupational aspiration. However, they are not in agreement with those of Strong who found that the correlation between OL scores and intelligence is .05. When intelligence is held constant, the correlation between OL scores and academic achievement is positive, but no longer significant. In the multiple correlation of academic achievement to the OL scores and intelligence, the correlation coefficient increases slightly. The multiple and partial correlations reveal that intelligence is the most important factor in academic achievement. The relationship between the OL scores and the Extracurricular Activity Index does not appear to be clearly significant or close.

Conclusions

The purpose of this study was to determine the degree of relationship which exists between the OL score of the Strong Vocational Interest Blank and various other presumed measures of level of aspiration, such as teachers' ratings of students' level of aspiration, students' self-ratings of level of aspiration, and two laboratory tests of level of aspiration. Secondly, an attempt was made to ascertain whether any relationship exists between OL scores and scholastic and extracurricular achievement. The experimental group consisted of 132 white males who were in the junior and senior classes of two high schools about 30 miles from New York City.

The results of this study seem to warrant the following conclusions:

1. The OL score (Strong's Vocational Interest Blank) is positively but not clearly significantly related to teachers' ratings of students' level of aspiration.

2. No significant relationships exist between the OL scores and the several laboratory tests of level of aspiration. However, there is some degree of relationship between OL scores and the other measures of level of aspiration.

3. There is a significant relationship between OL scores and academic achievement, but when intelligence is held constant, the coefficient of correlation is positive but no longer significant. In a multiple correlation of academic achievement with intelligence and OL scores, the correlation coefficient increases slightly. It appears that the OL score is not useful as a predictive tool for academic achievement. However, the data reveal that an individual's OL score is related to his intelligence and academic achievement. How the individual appraises his abilities and achievement appears to determine to a great extent the level of his choice of occupational goal.

4. There is no clearly significant relationship between OL scores and the extent of student participation in extracurricular activities.

5. Although there are differences in the population, it has been found that OL scores are positively related to intelligence.

Level of aspiration is highly specific. One cannot infer from the data in Tables 3 and 4, not all of which have been discussed in this paper, the presence of a general trait "level of aspiration." The conduct of an individual is due to factors common to the situations represented in the test. Actions in a wide range of situations cannot be predicted from knowledge in one type of situation, since the varied situations are not sufficiently similar to elicit a common response from the individual.

The OL scale of Strong's Vocational Interest Blank is not significantly related to the other measures of level of aspiration, and is of little value as a predictive instrument of academic or extracurricular achievement when intelligence is held constant. The interpretation of the OL scale as a measure of level of aspiration is apparently not warranted. The OL score is so constructed that it should indicate the socioeconomic level of an individual's interests, but there is no evidence to suggest that it measures his level of aspiration or drive.

CHAPTER IV

THE RELATIONSHIP OF CERTAIN SOCIAL FACTORS TO OCCUPATIONAL LEVEL AS MEASURED BY STRONG'S VOCATIONAL INTEREST BLANK[*]

LAWRENCE H. STEWART

THIS study is concerned with the relationships between certain social factors presumably operative during early childhood and adolescence and the individual's level of vocational interests as measured by the OL scale of the Strong Blank.

Hypothesis. Occupational level as measured by Strong's Vocational Interest Blank is, in part, a function of social background. Sons of skilled workmen whose social backgrounds vary from that to be expected on the basis of fathers' occupations tend to make score[illegible] the OL key which vary from the mea[illegible] skilled group.

Problem and Method

Subjects. The subjects for the study were 136 high-school juniors and seniors who were sons of skilled workmen, skill being classified according to the *Dictionary of Occupational Titles* (57). All of the subjects were selected from five schools within commuting distance of, but not within, New York City. An attempt was made to get a complete coverage of each school. The subjects were sons of white parents born in the United States. Their ages ranged from 15 to 21 years with a mean age of 16.7 years.

Sims (44) has found that socioeconomic status is a relative term and is significant only in the group within which the child lives. Comparisons are most reliable when made within the same or similar groups. With this is mind, Newark, N.J. was selected as a location likely to provide the desired number and variety of subjects. However, examination of data on the subjects obtained at the West Side High School in Newark showed that the proportion of Italian students was unduly large. In order to reduce the possibility of getting too many subjects of one national origin, other neighboring communities were selected as a source for the remaining subjects.

Sons of skilled workmen were selected for study, since the mean OL standard score of skilled workmen is approximately 50. This permitted the scores of the subjects to deviate in either direction from the mean score of the skilled group. If either the professional or unskilled occupational group had been chosen, the OL scores of the sons could have deviated in one direction only.

Satisfaction with the job depends, in part, on the level of the job in the manual or white-collar hierarchies (51). Skilled workers are more satisfied with their jobs than semiskilled workmen; professional workers more satisfied than clerical workers. Davidson and Anderson (18) have shown that clerical workers are the main reservoir for the higher white-collar occupations while the semiskilled and unskilled workers are the reservoirs for the skilled occupations. Choosing sons of skilled workmen, who have reached the top of the manual hierarchy, may have partially eliminated the effects of parental dissatisfaction and the desire of the parents to move upward in the occupational world.

Colored students and sons of foreign-born parents were excluded from the sample, because it was not known what effects race or foreign cultures would have on the variables being investigated.

---

[*] Part of an Ed.D. investigation conducted at Teachers College, Columbia University, under the direction of Dr. Donald E. Super, Dr. Nicholas Hobbs, and Dr. Herbert Zucker.

Instruments. The OL scale of the Strong Blank, a form of the Otis Test of Mental Ability, and a questionnaire designed to elicit information concerning the social background of the individual were used in the study. The first two instruments have been described elsewhere (26, 49, 53). The questionnaire consists of 49 items exclusive of identifying data (Appendix D). It can be divided into two parts: items concerning the subjects' family background, and those concerning the subjects' activities either in or outside the home.

Administering the instruments. The Strong Blank and the questionnaire were given to groups of from five to ten subjects. In order to motivate the subjects, they were informed that the Strong scores would be made available to them through their school counselor. After five minutes practice on the Strong, they were told to complete the blank outside of school, and attention was turned to the questionnaire, which was completed in the group. The use of small groups permitted the investigator to answer questions and to check each questionnaire for omissions and ambiguous answers. Either the Otis Quick-Scoring (Gamma) or Self-Administering (Higher) IQ's were taken from school records. The school administered the Otis Quick-Scoring Test to a few students for whom Otis IQ's were not available.

The use of two forms of the Otis tests raised the problem of comparability of IQ's. Information furnished by the publisher indicates that the Quick-Scoring averages about one raw score higher than the Self-Administering tests. Since one raw score is roughly equivalent to one IQ point, Self-Administering IQ's were converted to Quick-Scoring IQ's by the addition of one IQ point.

Statistical analysis. The analysis of variance was used as one variable was in the form of multiple categories and the other in continuous standard scores.

Analysis of the data indicated that OL scores were not related to school or grade or clearly related to age. Therefore all of the subjects were considered as a single sample.

Findings are summarized in Table 5, in which eight significant relationships are reported (three of which are family data and five personal), together with 19 variables for which no relationship was found.

Discussion

The findings indicate that the mother may have a greater influence on the development of vocational interests than has [illegible]erto been assumed. Psychoana[illegible]ory stresses the tendency for the so[illegible]emulate the father, a fact which see[illegible]o be at variance with the above findings. However, Strecker (48) points out that sons tend to be emotionally attached to their mothers. Such a viewpoint provides a logical framework for these findings, which suggest the following theory.

The mother plays an important role in transmitting the development of cultural values and mores to the children, and therefore helps set the level of the son's vocational interests. Not only does the mother nourish the child physically, but she also provides a cultural atmosphere, highly saturated with her own interests and values, which is influential in determining those of the son. This foundation, established during childhood, controls the interpretation of future experiences and, to a great extent, the formation of interest patterns. The father may help determine the *vocational field*

TABLE 5

Summary of Findings

Related to OL Score	Not Related to OL Score
Significant at the .05 level: *Family* 1. Mothers' socioeconomic status as measured by: Occupation of maternal grandfather Occupation of maternal grandmother 2. Religion of the family *Subject* 3. Participation in high-school organizations 4. Type of summer work experience 5. Occupation expected after completion of education Significant at the .01 level: 6. Vocational and educational plans after completion of high school 7. Intelligence	*Family* 1. Occupational status of: Members of immediate family Paternal grandparents Most successful relative Person most admired 2. Educational level of relatives 3. Ethnic origin 4. Parents' birthplace 5. Financial status of friends of the family 6. Type of dwelling 7. Feelings concerning the neighborhood 8. Parents' participation in organizations 9. Supplementary income 10. Possessions at home *Subject* 11. Participation in out-of-school organizations 12. Holding offices in any type of organization 13. Source of spending money 14. After-school work experience 15. Travel 16. Visits to museums 17. Socioeconomic status of friends 18. Long-range vocational aspirations 19. Parental aspirations for the subject

of interest (problem discussed in Chapter I); the mother is seen as a determiner of the *socioeconomic level* of interest.

The other observed relationships, interpreted in light of this hypothesis, could be thought of as being reflections of this earlier training. Adolescents enter into extracurricular activities, and have vocational and educational aspirations, partly because of their earlier conditioning; they have developed interests which make these activities and aspirations compatible.

Very little evidence is available to support the above hypotheses, other than the findings of the present study, and these findings are by no means clear-cut. They do, however, point up the need for further research concerning the mother's role in the development of the level of the son's vocational interests.

Conclusions

The purpose of the study was to investigate the relationship between certain social factors, presumably operative during early childhood and adolescence, and occupational level as measured by Strong's Vocational Interest Blank.

A questionnaire designed to elicit social background information, the Strong Vocational Interest Blank, and a form of the Otis Tests of Mental Ability were administered to 136 high-school juniors and seniors who were sons of skilled workmen, classified according to the *Dictionary of Occupational Titles*. The subjects were sons of white parents born in the United States. All of the subjects were in schools surrounding metropolitan New York City.

When a large number of variables are investigated, a few seemingly significant relationships may be expected on the basis of chance. However, eight significant relationships out of 57 variables are more than the expected number of chance relationships.

It seems difficult logically to account for the fact that some of the variables are related to OL scores while other seemingly very similar factors are not related. This suggests the possibility that some of the quantitative information elicited by the questionnaire may not get at important aspects of the problem.

The following tentative conclusions seem justified:

1. The OL scores of sons of skilled workmen are *not* related to discrepancies in the socioeconomic status of the immediate family as indicated by occupation of siblings and mother, educational level of the members of the family, possessions at home, supplementary income, activities engaged in by parents, geographic mobility of parents, ethnic origin, evidence of social mobility such as feelings toward neighborhood and socioeconomic status of friends, and type of dwelling.

2. The OL scores of sons of skilled workmen *are* related to the mothers' socioeconomic status as determined by the occupational level of both maternal grandparents. These relationships, however, are not clearly significant.

3. The OL scores of sons of skilled workmen are *not* related to the educational level of maternal grandparents, paternal grandparents, and most educated relative; nor to the occupational level of paternal grandparents, most successful relative, and person most admired.

4. The OL scores of sons of skilled workmen *are* related to the religion of the family, although the relationship is not clear-cut. Sons of mixed marriages score significantly higher than sons of Catholics. Sons of Protestants score higher than sons of Catholics, but the difference is not quite statistically significant.

5. The OL scores of sons of skilled workmen *are* related to activities such as type of summer work experience and belonging to school organizations. The relationships, however, are not clearly significant.

6. The OL scores of sons of skilled workmen are *not* related to factors in the subjects' environment such as membership in out-of-school organizations, source of spending money, after-school work experience, travel, visits to museums, and socioeconomic status of closest friends.

7. The OL scores of sons of skilled workmen *are* clearly related to the indivi[illegible] immediate vocational or educational plans. Those expecting to go to college score significantly higher than those who plan to get a job on leaving high school, go to a business school, or go to a technical institute.

8. The OL scores of sons of skilled workmen *are* related to the type of job expected after completion of education. The relationship, however, is not clearly significant.

9. The OL scores of sons of skilled workmen are *not* related to the subjects' long-range vocational aspirations nor to the vocational and educational aspirations which they think their parents have for them.

10. The OL scores of sons of skilled workmen *are* related to intelligence. The correlation is low but clearly significant.

CHAPTER V

THE MEANING OF OCCUPATIONAL INTEREST LEVEL AS MEASURED BY STRONG'S BLANK

DONALD E. SUPER

In launching the series of investigations reported in the preceding chapters we had in mind questions such as the following:

> Is having interests like those of professional men a sign of motivation to strive to rise to the higher socioeconomic levels?
>
> Are adolescents with interests like those of business executives inclined to show, in their school work and activities and in their educational and vocational plans, patterns of behavior which can be interpreted as indications of high aspirational level?
>
> Are men whose level of occupational interests is high likely to be dissatisfied with a low level of vocational attainment, while those whose occupational interest level is [illegible] tend to be satisfied with unemployment, [illegible], and dependency?
>
> Is occupational interest level simply a [illegible]sure of the extent to which an individual has developed interests which characterize one or another of the various socioeconomic levels?
>
> With what factors and experiences in and outside of the home are differences in occupational interest levels associated?

The answers to these questions were sought in a review of previous studies of the OL scale of Strong's Vocational Interest Blank (the results of which have been omitted from this monograph), and in three investigations conducted with high-school boys and adults, in which OL scores were related to educational achievement, extracurricular and out-of-school activities, laboratory tests of level of aspiration, ratings of aspirational level, educational and vocational plans, congruence of vocational aspiration and attainment (as shown by satisfaction with unemployment), intelligence, and social background and status. In the following pages an attempt is made to synthesize these findings and to draw some conclusions as to the meaning and use of the OL scale of Strong's Blank.

Occupational Interest Level and Indices of Drive

Educational achievement. When intelligence is held constant, there is no significant relationship between the OL scores of high-school students and their grades, as shown in studies by Ostrom (43) and by Handelsman (27). Although the latter investigator found a clearly significant and moderately high intercorrelation before intelligence was partialled out, this relationship shrank to insignificance once mental ability was controlled.

At the college level, earlier studies by Kendall (33) and by Ostrom (42) have reported findings which lead the present writer to conclude that the relationship between occupational interest level and achievement in college is very slight. It was only when Ostrom compared the honor-point ratios of High-OL and High-IQ students with those of five other possible combinations of OL scores and IQ's that clearly significant differences in college achievement were found. Working with graduate students of business, Strong (49) found no significant differences in occupational interest level between high and low achieving groups.

In the case of unemployed adults, Barnett has reported a low but possibly significant correlation ($P = .05$) between occupational interest level and amount of education, which can also be largely explained by the correlation of both of these with intelligence.

From the above data it seems necessary to conclude that there may be a very slight relationship between occupational interest level (as measured by Strong's Vocational Interest Blank) and educational achievement. This relationship is, however, not clear enough to justify concluding that occupational interest level is related to motivation to achieve. Neither is the relationship close enough to warrant using the scale as an educational counseling or selection instrument.

Extracurricular and out-of-school activities. Another possible way in which motivation to achieve or to lead might manifest itself is in the relatively free activities of school and community life. Handelsman found what might have been a significant but slight relationship between occupational interest level and membership and office-holding in school activities in one of the high schools, but not in the other. Stewart's group of high-school boys from homes of skilled fathers showed a relationship like that of Handelsman's first group when participation in in-school activities was analyzed, but none when occupational interest level was related to school office-holding, or to out-of-school activities, including part-time jobs. On the other hand, Stewart did find that the boys who engaged in professional types of occupations during the summer tended to have higher OL scores.

These results suggest that there is some tendency for persons with high occupational interest levels to seek outlets in school activities and to use their summer vacations for earning, orientation, or training, although they show no evidence of being inclined to hold school offices, to be active in organizations outside of school nor to work part time while in school (the last two are probably less rewarding psychologically and socially). This might be characterized as a middle-class pattern. In the case of Stewart's sons of skilled workers, it raises a question as to whether we have demonstrated a motivation to rise above the way of life of the paternal socioeconomic level, or simply the transmittal of a different way of life to the son by influences other than the father. As Stewart found a significant relationship between occupational interest level and the socioeconomic level of the mother's family among these sons of skilled workers, it seems probable that the mother has transmitted to the son values from a cultural level different from that of the father. Whether the interest in middle-class activities and the acceptance of middle-class values which result in participation in school activities and [illegible] getting summer jobs is the result [illegible] passive acceptance of a way of life or [illegible] effect of unusual motivation is not made clear by the data.

We are probably justified in concluding that occupational interest level is slightly related to the use of school and summer work opportunities in ways which contribute to the attainment of status and to the preservation of other middle-class values. Taken alone, this relationship is, however, so slight as to provide little basis for using the OL scale of Strong's Blank in counseling or selection.

Ratings of drive. In a study reported by Ostrom (43), students in the last year of high school were rated for drive by their teachers, their fellow students, and by the investigator after an interview. These judgments were based on the behavior of the students in the classroom, in extracurricular activities, and in a variety of other life situations. Ostrom found that all of these ratings were correlated with scores on the OL scale of Strong's Blank, and concluded that occupational interest level is a measure of motivation. He did not, however, allow for the fact that students from higher socioeconomic levels participate in school work and activities more than do those of lower socioeconomic levels, and that, as Hollingshead has shown (30), boys and girls from lower levels who want to participate often do not do so because they feel that they are discriminated against. Ostrom's results could therefore be spurious.

Our results tend to agree with Ostrom's, but show a much lower and much less clear-cut relationship. Handelsman, we have seen, found that teachers' ratings of pupil motivation had correlations of .23 and .25 with OL scores, significant at the .05 level, whereas in only [illegible]e of his two groups were student self-r[illegible]s related to occupational interest l[illegible]

In view of Hollingshead's (30) findings concerning teachers' attitudes toward lower-class children, it is quite possible that the low positive relationship which seems to exist between teacher ratings of motivation and occupational interest level is the result of a tendency on the part of teachers to rate boys according to the extent of agreement between the boys' apparent values and their own middle-class values. If so, Ostrom's and Handelsman's findings on ratings (whether by teachers, peers, or the investigator) and OL may only indicate that the boys whose interests are most like those of the middle-class majority in high school win the most social approval. If so, then occupational interest level scores could be considered measures of status rather than of drive. In using them as measures of motivation or drive on the basis of these findings one would therefore need to make assumptions the validity of which is open to question.

Laboratory measures of aspiration level. Handelsman's study shows that there is no relationship between standard laboratory tests of aspiration level and occupational interest level. Indeed, his findings appear to demonstrate that Frank's conclusion that level of aspiration is a fairly stable characteristic of personality must not be applied too broadly, for there is no relationship between the two laboratory measures of aspiration level used by Handelsman and the other, more nearly lifelike, measures which he used. By contrast, these other (nonlaboratory) measures did in some instances yield some low but positive intercorrelations.

Educational and vocational plans. The nature of a student's plans for further education and subsequent vocation may be indices of level of aspiration which are more lifelike than those provided by the traditional laboratory experiments in aspiration level. Stewart has shown that occupational interest level is related to high-school seniors' plans for further education or work immediately after finishing high school, when paternal occupational level is held constant. Again, however, a positive finding must be qualified in view of the fact that Stewart found significant relationships between occupational level of maternal grandparents and OL scores: as sons of skilled workmen whose mothers come from superior socioeconomic backgrounds tend to have interests characteristic of higher occupational levels, perhaps they acquire also from their mothers a more positive valuation of post-high-school education and of white-collar jobs than is typical of boys whose mothers as well as fathers come from skilled or semiskilled backgrounds.

Operationally speaking, occupational interest level is a measure of similarity of interests to those of men at a given socioeconomic level. If it is also an index of motivation to rise to higher social levels, other evidence must be found to that effect.

The picture is further complicated by the fact that Stewart found no relation between occupational interest level and long-range vocational plans. It may be that the long-range vocational plans of these high-school seniors are unrealistic; there is ample evidence to this effect for other groups in the literature (45), and disparities between aspirations and expectancy are greater at the lower socioeconomic levels than at the higher. If this is so, these data cannot provide a criterion for the OL scale.

Vocational achievement. The first two investigations of the relationship between scores on the OL scale of Strong's Vocational Interest Blank were made by Strong himself (49). In a study of Forest Service employees, the higher job levels were characterized by significantly higher mean OL scores, which resulted from the relative absence of men with low occupational interest levels in the administrative groups. Furthermore, when a group of unemployed men were studied, 25 per cent were found to have occupational interests like those of unskilled and semiskilled workers. These findings were viewed by Strong as suggesting that occupational interest level might be indicative of drive to achieve status.

Barnett also studied unemployed men. Although his subjects were tested during a boom period and Strong's in a depression, Barnett's chronically unemployed group made the same average OL score as Strong's; his nonchronic group appropriately scored about one-half sigma higher. Most important of all, strikingly high negative correlations between occupational interest level and satisfaction with being unemployed were found for both groups (−.73 and −.68), and these were not lowered when intelligence was held constant. Taken as they stand, these results appear to offer a clear-cut confirmation of the hypothesis that occupational interest level as measured by Strong's Blank is an index of drive: men who are satisfied with being unemployed tend to make low OL scores, while those who are dissatisfied with their lowly status tend to make high scores.

But if occupational interest level is indicative of drive, why have these normally intelligent, normally educated, and economically favored men (witness the Otis scores, educational levels attained, and [illegible]imed parental occupations) not [illegible] more successful? Why are the many hig[illegible]L's who are dissatisfied with their status unemployed and dependent in a period of labor shortages? In some instances (or perhaps many) still other factors may be related to their being unemployed: physical handicaps were ruled out by the sampling methods, but personality maladjustments are an obvious possibility.

But Barnett reports some other findings which bear on this problem. Whereas parental occupation *was not* related to occupational interest level in his subjects, parental occupation *was* negatively related to being satisfied with unemployment. In other words, although there is no relation (in this group) between father's occupation and son's occupational interest level, both of these factors are significantly and negatively related to being satisfied with unemployment. *These findings suggest that the OL scale measures what it was constructed to meas-*

ure: the socioeconomic level at which a person's interests would place him. Men who are unemployed are unable financially to satisfy the interests they have developed in childhood and adolescence; the higher the social scale placement of their interests the less able they are to satisfy these interests when unemployed and the more dissatisfied they are with their unemployed status. Similarly, men whose fathers were employed at relatively high socioeconomic levels find being unemployed more in conflict with their socially acquired values than do men whose fathers were from lower occupational levels and who therefore were more inclined to consider some unemployment inevitable. For these reasons occupational interest level is related to unemployment satisfaction in the same way as is parental occupational level.

If this is indeed the nature of the relationships, there is no need to postulate that motivation or drive is involved in occupational interest level. Occupational interest level may be *static* in nature, simply an indication of the status level of interests, or it may be *dynamic* in the sense of moving one to do something about finding appropriate outlets for his interests. While it is true that data such as Strong's follow-up studies of Stanford students (49) suggest that there is something dynamic in interests (Strong's studies reveal a tendency to enter occupations in which interest is high or, if a low-interest occupation is entered, to shift to fields of greater interest), the dynamics in this instance are those of *fields,* rather than of *levels.* And it is in level that we are interested in this context.

We must conclude, then, that it has not as yet been demonstrated that occupational interest level is a measure of motivation to achieve or to acquire status. It is, at least, a measure of similarity of interests to those of men at various socioeconomic levels (its method of construction should guarantee this). The Forest Service administrators studied by Strong may have been attracted to their higher-level jobs by the administrative nature of the work rather than by desire for status, and the forest rangers may have remained rangers not because they lacked ambition but because they liked a ranger's work more than management. True, the OL scale may yet be shown to measure drive, and we have not demonstrated that it does not do so. But we have shown that all available evidence concerning the OL scale can parsimoniously be interpreted in terms of *similarity of interest.* It is not unreasonable to expect a measure of similarity of interest to measure only similarity of interest. Neither is it surprising to find that *fields* of interest have a dynamic which may not be shared by *levels* of interest, when the instrument used in measuring them both was originally designed to assess the former.

In counseling and selection, then, it would appear justifiable to use the OL scale in order to find out whether a person has interests which will find appropriate outlets at a given occupational level, or in order to ascertain at which occupational level he is most likely to find appropriate outlets for his interests. But so far as we now know *level* of interests is likely to be less important to adjustment than is employment in a field of work which provides appropriate outlets for types of interests.

Occupational Interest Level and Personal Data

The data and conclusions so far reviewed have had to do with relations between occupational interest level and various indices of drive, whether tested, rated, or manifested in school, leisure-time, or vocational activities. However, in order to understand the meaning of occupational interest level we must also know something of its origins, as deduced from its relationships with various types of personal and family data.

Intelligence. In the original studies of the OL scale, Strong (49) found no relation between scores on that scale and intelligence test scores. Our investigations contradict the original work, with correlations ranging from .20 to .42, both of the latter being for Handelsman's high-school juniors and seniors; Stewart's high-school seniors, all of them sons of skilled workmen, yielded a correlation of .25 (significant at the 1 per cent level), and Barnett's men an *r* of .28. It therefore seems legitimate to conclude that in a typical group there is a relationship between occupational interest level and intelligence. One must understand at least a little in order to be interested.

Social background. No previous investigations have dealt with this topic, on which Barnett gathered a little data and on which Stewart's study focused. Barnett related OL scores to paternal occupation in a socially, educationally, and intellectually heterogeneous group of unemployed men, finding no relationship in the chronic group (.14) and a very low but significant negative correlation in the nonchronic (−.24). These are surprising results, as one might have anticipated that the social level of the father would be a prime determinant of the social level of the son's vocational interests. However, the correlation between *field* of interests in father and son has been shown to be only .30 (6, 49).

On the other hand, Stewart did find a positive and significant relationship between the occupational interest level of his skilled workmen's sons and the occupational levels of both maternal grandparents. As paternal occupation was held constant in this study we may not conclude that the role of the mother is more important than that of the father, but the finding does fit in with the widely held idea that women are the carriers of culture. It suggests that when the mother's background is superior or inferior to that of the father, the interests and values which typify her background will to some extent be transmitted by her to the son.

Only one other item of family background was found to be related to occupational interest level: family religion. Sons of mixed Protestant-Catholic marriages tended, in Stewart's sample, to make higher OL scores than others, Protestants ranked next, and Catholics [illegible] (the difference between the last two groups was not statistically significant and there were too few Jews for comparisons). Perhaps willingness to marry across religious lines is the result of a social mobility which manifests itself in higher vocational *aspirations* in the children of such marriages, but it may also be that this freedom from tradition is associated with interests and values which affect the upbringing of the children and manifest themselves in the children's interests. Perhaps only the more highly educated are likely to feel free to marry across religious lines.

Other family data analyzed by Stewart included the occupations of mother and siblings, the educational attainments of other members of the family, parental community activities, parental social mobility as indicated by attitudes toward the neighborhood in which they live and by the social levels of people associated with, and similar factors which might be expected to provide models of attainment and aspiration for maturing children. None of these, however, was related to occupational interest level.

These findings are not easily synthesized or rationalized. If we explain away the lack of significance of the mother's occupation on the grounds that women's occupations mean little in terms of status because of limitations on their employment (hardly true when one examines the social backgrounds of teachers, nurses, stenographers, sales clerks, factory operatives, and domestics, as shown by Hollingshead), we must then explain the statistically significant relationship with the maternal grandmother's occupation. And why is occupation more important than education? It might be because the former clearly provides a status label, whereas the latter does not—but then the occupations of the most successful relative and of siblings should be import[illegible] to occupational interest level, whic[illegible]ey are not.

The most striking and clear-cut finding concerning family background and occupational interest level is the lack of relationship between the two. Although we are given a suggestion that perhaps the social level of the mother plays a part in determining the level of the son's occupational interests, this is not a conclusion of which we can be sure in view of seemingly contradictory findings. What, then, are the origins of interest, what the determinants of the occupational level to which a person's interests correspond?

Perhaps *field* or type of interest is the important factor, the *level* factor being important only in that it reflects the level at which the individual's type of interest can find outlets. Clerks, skilled workmen, and semiskilled laborers do not manage enterprises, hence a person interested in managerial activities and able to do so finds outlets in administrative work; *ipso facto,* he has interests like those of men at high occupational levels. Engineers do not usually work directly with tools and materials, therefore a man with technical interests who likes to work directly with materials has, by virtue of the way in which work is now organized, interests like those of skilled workmen. Empirical support for this hypothesis is reported in Table 6.

In Table 6 data from Strong (49) have been reworked to show the mean OL scores of the various types of work. Groups 1 through 8 represent *fields* of work and to a lesser extent, perhaps, levels, while Groups 9 and 10 are a more nearly equal mixture of field and level. As the mean OL scores show, persons who have literary interests are *ipso facto* likely to have a high occupational interest level. Similarly, those who are interested in skilled work with tools are *ipso facto* likely to be placed at a low occupational level.

If this is correct, it may be necessary to seek the origins of occupational interest levels in the origins of types or fields of interests rather than in status or prestige drives. The origins of types of interests have been dealt with at length elsewhere (53, Ch. 16).

General Conclusions

In conclusion, it has been demonstrated that the Occupational (Interest) Level scale of Strong's Vocational Interest Blank is not related to family background, although the socioeconomic level of the mother's family and the freedom from binding traditions represented by willingness to marry across religious lines may have some bearing on occupational interest level. Occupational interest level

TABLE 6

Occupational Field and Occupational Level
(Data reworked from Strong (49), Tables 50, 53, and 87)

Field or Occupational Family	N	Occupational Level		r, Field × Level[1]
		M	SD	
1. President, Mfg. Concern	169	63.4	7.2	.63
2. Prof.: Literary-Legal	628	63.2	6.7	.53*
3. Biological Sci.	764	60.0	7.1	−.13*
4. Physical Sci.	1083	60.9	6.7	−.22*
5. Social Welfare	882	58.9	7.2	−.17*
6. Business Contact	786	62.0	7.5	.43*
7. Business Detail	1128	57.7	7.4	−.14*
8. Skilled	—	48.3	8.0	—
9. Semiskilled	—	45.9	8.6	—
10. Unskilled	—	44.4	8.7	—

* Pearsonian r's for each *occupation* in the family in question averaged by z transformation to yield an average r for the occupations in each family.

[1] This is the correlation between scores on the occupational scale (representing field of interest) and scores on the OL scale. The populations in each case include a wide range of occupational field scores, skewed toward the upper end of the range.

is not simply a reflection of the father's occupational level. It seems most likely that occupational interest level is a function of the occupational levels at which various fields or types of interests can manifest themselves, and hence indirectly of ability, aptitudes, personality, and social evaluation.

Occupational interest level is not closely or clearly related to educational achievement. It is slightly related to extent of participation of extracurricular activities and summer employment on the part of high-school students, although not to school office-holding; these facts suggest that OL scores reflect middle-class values and interests. Occupational interest level is somewhat related to ratings of drive made by teachers, peers, and observers, but this may simply be a manifestation of the way in which middle-class people tend to approve of those whose values most resemble their own. It is not related to laboratory measures of level of aspiration. It is related to immediate post-school plans of high-school students, but this may reflect the inherent relation between field of interest and level, more than the effect of occupational level aspirations upon choice of field. Although the method of construction of the OL scale r[illegible]s in a relationship between occupati[illegible] interest level scores and occupation[illegible] placement on the socioeconomic ladder, there is no evidence which would justify concluding that the OL scale, is a measure of drive to advance occupationally: it is, however, clearly at least a measure of similarity of interests to those of men at various points of the socioeconomic scale. *The evidence now available warrants interpreting the occupational level score as a measure of status of interests:* the OL score can indicate the socioeconomic level at which a person should be able to find outlets for his interests. *The evidence does not warrant interpreting the OL score as a measure of drive:* it has not been demonstrated that the OL score can or cannot indicate how much a person will exert himself to rise on the occupational ladder or to succeed in his field of activity.

APPENDIX A

UNEMPLOYMENT SATISFACTION SCALE

GORDON J. BARNETT (after Robert Hoppock)

A. Choose the one of the following statements below that you think best tells your feelings about being unemployed. MARK IT WITH AN X.

——— I hate it.
——— I dislike it.
——— On the whole I don't like it.
——— I am indifferent to it.
——— On the whole I like it.
——— I like it a good deal.
——— I like it very much.
——— I am enthusiastic about it.
——— I love it.

B. Mark an X in front of the statement below that best tells how you think you compare with other unemployed people.

——— No one likes being unemployed better than I do.
——— I like being unemployed much better than most other unemployed people.
——— I do not know how I compare with other people.
——— I dislike being unemployed more than most unemployed people.
——— I dislike being unemployed *much* more than most unemployed people.
——— No one dislikes being unemployed more than I do.

C. Mark an X in front of the statement below that tells HOW MUCH OF THE TIME you feel SATISFIED with being unemployed.

——— All of the time.
——— Most of the time.
——— A good deal of the time.
——— About half of the time.
——— Occasionally.
——— Seldom.
——— Never.

D. Mark an X in front of the statement below which best tells how you feel about changing your unemployed condition.

——— I would give anything to stop being unemployed.
——— I would take a job if I could find anything to do.
——— I would take a job if it bettered my condition.
——— I would not want to stop being unemployed altogether.
——— I cannot think of any job for which I would stop being unemployed.
——— I would not give up being unemployed for any job.

APPENDIX B

STUDENT'S SELF RATING

IRVING HANDELSMAN

Student's Name .... Date ....

Please check the sentence which best describes you.

( ) I believe that I set high goals for myself, and work hard to attain them.
( ) I believe that I set high goals for myself, but do not do much about them.
( ) I believe that I set average goals for myself, and work hard to attain them.
( ) I believe that I set average goals for myself, but do not do much about them.
( ) I set no goals for myself.

APPENDIX C

RATING OF LEVEL OF ASPIRATION

IRVING HANDELSMAN

Student's Name .... Date ....

Please check the sentence which most describes this student:

( ) He sets high goals for himself, which he works hard to attain.
( ) He sets high goals for himself, which he does not work hard to attain.
( ) He sets average goals for himself, which he works hard to attain.
( ) He sets average goals for himself, which he does not work hard to attain.
( ) He sets no goals for himself.

....
Teacher's Name

APPENDIX D

BIOGRAPHICAL QUESTIONNAIRE

LAWRENCE H. STEWART

Name .... Age ....
School .... Grade ....

1. How many older brothers do you have? .... Younger? ....
2. How many older sisters do you have? .... Younger? ....
3. Check any of the following statements that apply to [illegible]
   ( ) I live with my parents.
   ( ) My parents are divorced and I live with one of them.
   ( ) My father is dead.
   ( ) My mother is dead.
   ( ) I live with foster parents (or guardian, grandparent, etc.); (fill in date) since 19.......
4. Father's (or foster father's stepfather's) job (for example: baker or carpenter): ....
   ....
   Place of work (for example: flour mill or machine shop): ....
   ....
   Just what does he do? ....
   ....
5. Mother's job (for example: dressmaker or housewife): ....
   Place of work (if she works away from home): ....
   Just what does she do? ....
   ....
6. Most successful older brother's job: ....
   Place of work: ....
   Just what does he do? ....
   ....

7. Most successful older sister's job: ....
Place of work: ....
Just what does she do? ....
....

8. Father's father's job: ....
Place of work: ....
Just what does or did he do? ....
....

9. Father's mother's job: ....
Place of work: ....
Just what does or did she do? ....
....

10. Mother's father's job: ....
Place of work: ....
Just what does or did she do? ....
....

11. Mother's mother's job: ....
Place of work: ....
Just what does or did she do? ....
....

12. Most successful relative's job: ....
Who is he or she (for example: cousin, uncle)? ....
Place of work: ....
Just what does he or she do? ....
....

13. Who is the person, living or dead, that you most admire (for example: aunt, local policeman, or John Paul Jones, naval hero)? ....
Just what does or did he do? ....
What are or were his major interests or outstanding characteristics? ....
....

14. Your parents have usually considered that they belong to the following religious groups (check one in each column):

Father	Mother	
( )	( )	Protestant
( )	( )	Roman Catholic
( )	( )	Jewish
( )	( )	Other (name) ....
( )	( )	None

15. Your parents and their parents are of (check one in each column):

Father	Mother	
( )	( )	Old American stock (your parents and grandparents born in this country).
( )	( )	Northern European stock (English, Scotch, Irish, Scandinavian, German, etc.).
( )	( )	Eastern European stock (Russian, Polish, Rumanian, Servian, Greek, Latvian, etc.).
( )	( )	Stock not indicated in the above groups.

16. Your parents were born in (check one in each column):

Father	Mother	
( )	( )	the South.
( )	( )	the West.
( )	( )	the Northeast.
( )	( )	the Midwest.

17. Your family's friends are (check one):
( ) less well off financially than you are.
( ) better off financially than you are.
( ) about the same as you are.

18. How far do you believe each of the following people went in school (place checks in the appropriate columns)?

	Grade School	Entered H.S.	Finished H.S.	Entered College	Finished College	Still in School	Don't Know
Father							
Mother							
Father's mother							
Father's father							
Mother's mother							
Mother's father							
Most educated brother							
Most educated sister							
Most educated relative (not mentioned above)							

Who is he (for example: uncle, cousin? ....

19. Your parents are active in (check one in each column):

Father	Mother	
( )	( )	no clubs or organizations.
( )	( )	one or two groups, generally just as a member.
( )	( )	a number of organizations, just as a member.
( )	( )	one or more organizations, often as an officer or leader.

List the principal organizations for:
your father ....
your mother ....

20. Do your parents go to concerts (check one)?
( ) Never
( ) Seldom
( ) Frequently

21. Your family lives in a (check one):
    - ( ) private, one-family house which they own.
    - ( ) private, one-family house which they rent.
    - ( ) two- or three-unit house or apartment which they rent.
    - ( ) two- or three-unit house or apartment which they own.

22. Your family feels that you (check one):
    - ( ) should live in a better neighborhood.
    - ( ) should live in a less expensive neighborhood.
    - ( ) are well located where you are at present and should continue the arrangement.

23. Approximately how many books does your family have (count about 25 books for every three feet of shelf)? ...... How many children's books? ...... How many adults' books? ......

24. Does your family subscribe to any magazines? Yes ______ No ______ If yes, list the magazines.
    ....

25. Does your family own any automobiles? Yes ______ No ______ If you own more than one automobile give the following information for each of them.

    Year .... Make .... Bought new? Yes ______ No ______

    Year .... Make .... Bought new? Yes ______ No ______

26. Do you have a telephone at home? Yes ______ No ______ If so, do you have a party line? Yes ______ No ______ Private line? Yes ______ No ______

27. Do you have a typewriter at home? Yes ______ No ______

28. Do you have a radio at home? Yes ______ [illegible] ______

29. Do you have a television set at home? Yes ______ No ______

30. Do you have a musical instrument at home? Yes ______ No ______ If yes, list the instrument or instruments. ....

31. Do your parents have an income other than that from their usual job (for example: rent from another house, interest on investments, etc.)? Yes ______ No ______

32. In school you belong to (check one):
    - ( ) no organizations.
    - ( ) one or two organizations.
    - ( ) more than two organizations.

33. If you indicated that you belonged to any organizations list them below.
    ....

34. In school you hold (check one):
    - ( ) no office in any of the above organizations.
    - ( ) an office in one or two of these organizations.
    - ( ) an office in more than two of these organizations.

35. Outside of school you belong to (check one):
    - ( ) no organizations.
    - ( ) one or two organizations.
    - ( ) more than two organizations.

36. If you indicated you belonged to any organizations list them below.
    ....

37. Outside of school you hold (check one):
   ( ) no office in any of the above organizations.
   ( ) one or two offices in the above organizations.
   ( ) more than two offices in these organizations.

38. Your closest friends are largely children of (check one):
   ( ) skilled workers and foremen (machinists, carpenters, etc.).
   ( ) proprietors and officials (managers, contractors, etc.).
   ( ) unskilled workers (janitors, street cleaners, etc.).
   ( ) semiskilled workers (filling station operators, lathe hands, etc.).
   ( ) professional men and women (lawyers, doctors, teachers, etc.).
   ( ) clerks and salesmen (bookkeepers, cashiers, mail clerks, etc.).

39. Check one of the following:
   ( ) I have an allowance.
   ( ) My family gives me what I need when I ask them for it.
   ( ) I have no allowance.
   ( ) I earn my own spending money for most of the smaller things I want.
   ( ) I earn all my spending money.

40. Do you usually work after school? Yes ______ No ______ If so, what was your last job? ....
....

For whom did you work (for example: grocer, baker, etc.)? ....

41. Do you usually work during the summer vacation? Yes ______ No ______

If yes, what was your last job? ....

For whom did you work (for example: grocer, ba[illegible]tc.)? ....

42. Have you ever traveled considerable distances from [illegible]? Yes ______ No ______

If the answer is yes, list your travels. ....
....

43. Do you visit museums (check one)?
   ( ) Never
   ( ) Seldom
   ( ) Frequently

44. What do you plan to do after you finish high school (check one)?
   ( ) Get a job.
   ( ) Go to a business school.
   ( ) Go to a college or professional school.
   ( ) Go to a technical institute.

45. What do you think your parents would like you to do when you finish high school (check one)?
   ( ) Get a job.
   ( ) Go to a business school.
   ( ) Go to a college or professional school.
   ( ) Go to a technical institute.

46. What kind of a job do you expect to get when you finish your education? ....

47. What kind of a job do you think you are likely to have ten years from now? ....

48. If you could have any job you wanted, as an adult, what would you like to do? ....
....

49. What kind of a job do you think your parents would like you to have? ....
....

BIBLIOGRAPHY

1. Ames, V. Factors related to high school achievement. *J. educ. Psychol.*, 1943, **34**, 229-236.
2. Barnett, G. J. *A Study of satisfied and dissatisfied chronically unemployed men.* Unpublished Ph.D. thesis, Teachers College, Columbia Univer., 1950.
3. Barr, F. E. Barr Scale Ratings of occupational status. In Terman, L. M., *Genetic studies of genius,* I. Palo Alto: Stanford Univer. Press, 1925.
4. Bell, H. M. *Youth tell their story.* Washington, D.C.: American Council on Education, 1938.
5. Berdie, R. F. Prediction of college satisfaction and achievement. *J. appl. Psychol.,* 1944, **28**, 239-245.
6. Berdie, R. F. Factors related to vocational interests. *Psychol. Bull.,* 1944, **41**, 137-157.
7. Bordin, E. S. A theory of vocational interests as dynamic phenomena. *Psychol. Measmt,* 1943, **3**, 49-65.
8. Bradley, W. A., Jr. Correlates of vocational preference. *Genet. Psychol. Monogr.,* 1943, **28**, 99-169.
9. Canning, L., *et al.* Permanence of voc[illegible]al interests of high school boys. *J[illegible]uc. Psychol.,* 1941, **32**, 481-494.
10. Carter, H. D. Twin similarities in occupational interests. *J. educ. Psychol.,* 1932, **23**, 641-655.
11. Carter, H. D. The development of vocational attitudes. *J. consult. Psychol.,* 1940, **4**, 189-191.
12. Carter, H. D. *Vocational interest and job orientation, a ten-year review.* Stanford: Stanford Univer. Press, 1944.
13. Carter, H. D. The development of interests in vocations. *Yearb. nat. Soc. Stud. Educ.,* 1944, **43**, 255-276.
14. Carter, H. D., & Jones, M. C. Vocational attitude patterns in high school students. *J. educ. Psychol.,* 1938, **29**, 321-334.
15. Carter, H. D., Taylor, K. V. F., & Canning, L. B. Vocational choices and interest test scores of high school students. *J. Psychol.,* 1941, **11**, 297-306.
16. Darley, J. G. A preliminary study of relations between attitude adjustment and vocational interest tests. *J. educ. Psychol.,* 1938, **29**, 467-473.
17. Darley, J. G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank.* New York: Psychological Corporation, 1941.
18. Davidson, P. E., & Anderson, H. D. *Occupational mobility in an American community.* Palo Alto: Stanford Univer. Press, 1937.
19. Dunkelberger, G. E. Do extra-curricular activities make for poor scholarship? *J. educ. Sociol.,* 1935, **9**, 215-218.
20. Frank, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.,* 1935, **47**, 119-128.
21. Frank, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.,* 1935, **47**, 285-293.
22. Frank, J. D. The influence of the level of performance in one task upon the level of aspiration in another. *J. exp. Psychol.,* 1935, **18**, 159-171.
23. Frank, J. D. A comparison between certain properties of the level of aspiration and random guessing. *Amer. J. Psychol.,* 1936, **3**, 43-62.
24. Frank, J. D. Level of aspiration test. In Murray, H. A., *et al. Explorations in personality.* New York: Oxford Univer. Press, 1938, 461-471.
25. Frank, J. D. Recent studies of the level of aspiration. *Psychol. Bull.,* 1941, **38**, 218-225.
26. Freeman, F. N. *Mental tests.* Cambridge: Houghton Mifflin, 1950.
27. Handelsman, I. *The relationship between Occupational Level scores and various measures of aspiration and achievement.* Unpublished Ed.D. Project, Teachers College, Columbia Univer., 1950.
28. Hartson, L. D. The validation of the rating scales used with candidates for admission to Oberlin College. *Sch. & Soc.,* 1932, **36**, 413-416.
29. Hartson, L. D. Further validation of the rating scales used with candidates for admission to Oberlin College. *Sch. & Soc.,* 1937, **46**, 150-160.
30. Hollingshead, A. B. *Elmtown's youth.* New York: John Wiley, 1949.
31. Hoppock, R. *Job satisfaction.* A publication of the national occupational conference. New York: Harper, 1935.
32. Jordaan, J. P. *Socio-economic status and the measured vocational interests of mechanically gifted boys.* Unpublished Ph.D. Thesis, Teachers College, Columbia Univer., 1949.
33. Kendall, W. E. The Occupational Level Scale of the Strong Vocational Interest Blank. *J. appl. Psychol.,* 1947, **31**, 283-288.

34. Kerr, W. A. The measurement of home environment and its relationship with certain other variables. *Purdue Univer. Stud. higher Educ.*, 1942, No. 15, 7-43.

35. Kitson, H. D. *The psychology of vocational adjustment.* Philadelphia: Lippincott, 1925.

36. Kitson, H. D. Creating vocational interest. *Occupations*, 1942, **20**, 567-571.

37. Lehman, H. C., & Witty, P. A. *The psychology of play activities.* New York: A. S. Barnes and Co., 1927.

38. Lewin, K., *et al.* Level of aspiration. In Hunt, J. McV., *Personality and behavior disorders*, I. New York: Ronald Press, 1944.

39. Mehus, O. M. Extra-curricular activities and academic achievement. *J. educ. Sociol.*, 1932, **6**, 143-149.

40. Mehus, O. M. Academic achievement of college students in different kinds of extracurricular activities. *J. educ. Sociol.*, 1934, **8**, 50-56.

41. Neel, M. O., & Mathews, C. O. Needs of superior students. *J. higher Educ.*, 1935, **6**, 29-34.

42. Ostrom, S. R. The OL key of the Strong Vocational Interest Blank for men and scholastic success at college freshmen level. *J. appl. Psychol.*, 1949, **33**, 51-54.

43. Ostrom, S. R. The OL Key of the Strong test and drive at the twelfth grade level. *J. appl. Psychol.*, 1949, **33**, 240-248.

44. Sims, V. M. *The measurement of socio-economic status.* Bloomington, Ill.: Public School Publishing Co., 1928.

45. Sparling, E. J. *Do college students choose vocations wisely?* New York: Bureau of Publications, Teachers College, Columbia Univer., Contributions to Education, No. 561, 1933.

46. Stewart, L. H. *The relationship of certain social factors to occupational level as measured by Strong's Vocational Interest Blank.* Unpublished Ed.D. Project, Teachers College, Columbia Univer., 1950.

47. Stewart, N. AGCT scores of army personnel grouped by occupation. *Occupations*, 1947, **26**, 5-41.

48. Strecker, E. A. *Their mothers' sons.* New York: Lippincott, 1946.

49. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford: Stanford Univer. Press, 1943.

50. Stubbins, J. The relationship between level of vocational aspiration and certain personal data. *Genet. Psychol. Monogr.*, 1950, **41**, 327-408.

51. Super, D. E. Occupational level and job satisfaction. *J. appl. Psychol.*, 1939, **23**, 547-564.

52. Super, D. E. *The dynamics of vocational adjustment.* New York: Harper, 1942.

53. Super, D. E. *Appraising vocational fitness.* New York: Harper, 1949.

54. Thorndike, E. L. *Adult interests.* New York: Macmillan, 1935.

55. [illegible]ney, A. H. Intelligence motivation and achievement. *J. educ. Psychol.*, 1931, **22**, [illegible]-434.

56. Umstattd, J. G. *Student self-support at the University of Minnesota.* Minneapolis: Univer. of Minnesota Press, 1932.

57. United States Employment Service, *Dictionary of occupational titles.* Washington, D.C.: Government Printing Office, 1939.

58. Wren, H. A. *Vocational aspirational levels of adults.* New York: Teachers College, Columbia Univ., Contributions to Education, No. 855, 1942.

VOLUME 66
NUMBER 11

WHOLE NO. 343
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Objective Measurement of Reality-Contact Weakness

By
ALBERT KREINHEDER
Los Angeles, California

Based on a Dissertation Presented to the General Faculty of the Claremont Graduate School in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Accepted for publication, January 17, 1952

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH ST. N.W., WASHINGTON 6, D.C.

Copyright, 1952, by the

American Psychological Association, Inc.

OBJECTIVE MEASUREMENT OF REALITY-CONTACT WEAKNESS[1]

I. *Introduction* 1
II. *Scoring methods* 7
III. *Results* 11
IV. *Implications of findings* 19
V. *Summary* 22
References 23

I. INTRODUCTION

The purpose of this study is to examine experimentally the view that maladjustment involves a degree of difference between what the maladjusted individual sees as real and what his society sees as real. The investigation of this viewpoint has entailed the construction of instruments for measuring perceptual deviations from objective or social reality and an examination of the inner consistency and predictive value of these instruments.

An assumption underlying the research is that perception cannot be entirely separated from memory, thought, and action—that to a large extent individuals act as they perceive (3, p. 367). The adaptive efforts of the individual toward his environment arise out of the total configuration of forces existing in the person-in-an-environment field. The overtly visible aspects of this configuration and the phenomenal aspects (perception) are at the instant of action in perfect relationship since they are but different ways of viewing the same events. From this point of view, therefore, behavior at odds with the demands of the objective environment coexists with perceptions that ignore, distort, or circumvent the real world as most people understand it.

A. Definition of Terms

Defined in terms of the procedures of the study, reality denotes those objectively apparent, verifiable relationships which are generally accepted as self-evident. Strength of the reality contact denotes, in turn, the degree to which an individual's perceptions conform to this normative standard. Strong reality contact is evident in any instance of accurate perception of objective stimuli or in instances of objectively oriented behavior. Weakness of the reality contact may be inferred when individuals are unable to perceive objective stimuli accurately—when, despite deliberate attempts at objectivity, experience and behavior are organized with reference to subjective factors and in disregard of objective fact. Experience, of course, is always subjective, and no weakness of reality contact is implied in the investing of objects with individual significance. Good reality contact merely demands that the raw facts and intractable relations of objective data be not ignored.

---

[1] The writer wishes to express his appreciation to the members of his doctoral committee for their helpful criticism and for the close association enjoyed with them: Professors F. Theodore Perkins, Florence Mateer, and Robert S. Ellis. He is also indebted to Dr. Bruce T. Jewell for performing the exacting task of ranking the 60 Rorschach records in order of adjustment, to Superintendent Otto L. Gericke of the Patton State Hospital, Patton, California, for the freedom granted in the use of hospital facilities, and to William O. Walcott of the Patton State Hospital Psychology Department for his helpfulness at many times.

It is assumed that the real world may be accurately experienced and that the level of accuracy of one's experience of the real world may be quantitatively described. Bold as such assumptions may appear, they are nothing more than corollaries to the definition of reality which is employed. Whereas absolute reality, if such exists, is nonverifiable, the reality dealt with is a matter of statistical averages and can be known. Objective reality becomes a reference point, and a definable distance exists between this reference point and the point representing personal reality. The shorter this distance, the greater the strength of the reality contact. Reality contact is held to be a quantitative feature of all behavior, cutting across all qualitative dimensions; and, as such, it may furnish a needed conceptual tool for assessing the degree of pathology in various behavior disorders.

B. The Problem and Method

The directing hypothesis motivating the present study may be stated as follows:

A determination of the degree of discrepancy between personal reality and objective reality yields a measure that is relevant to an individual's total adjustment.

Specific subquestions derived from this hypothesis have been put to experimental test. These are as follows:

1. Are the reliabilities of the instruments (the several tests of reality contact used in the study) sufficient to warrant further exploration of their value as measures of the strength of the reality contact?

2. Are the validities of the instruments sufficient to warrant further exploration of their value as measures of the reality contact?

3. Are the measures obtained by means of the instruments related to a criterion of the general efficiency of the personality adjustment?

Rorschach results and psychiatric diagnoses were used as criteria of the level of personality adjustment. To insure an adequate test of the questions there were two groups of subjects, the one group composed of paranoid schizophrenics who may be expected to show reality contact weakness in clear form, the other group consisting of normal and psychoneurotic subjects. The latter group, which is called the normal-neurotic group, was made as heterogeneous as possible with respect to level of adjustment. The object of this heterogeneity was the creation of an experimental situation that might reveal a relation between test scores and the level of adjustment within the normal-neurotic range. Analysis of results included intergroup comparison, analysis on the level of both groups combined, and analysis on the level of the normal-neurotic group.

C. Subjects

1. *The Normal-Neurotic Group*

The subjects comprising the normal-neurotic group were 30 in number and conformed to the following criteria:

1. Young adults of both sexes with ages from 20 to 40.

2. Superior intellectual potential as evidenced by vocational status, educational attainment, or the production of one or more Rorschach responses of high form level.

3. Persons who as a group were as heterogeneous as possible in the efficiency of personality adjustment.

4. The absence of clear-cut psychotic behavior at the time of testing or at any time in the individual's past.

The members of the normal-neurotic group were relatively homogeneous in social and economic background and in educational level. Fifteen of the subjects were college graduates and all but two of the remaining had attended college. Adjustment ratings made from the Rorschach by means of the Munroe revised check list (7) indicated that the attempt to secure heterogeneity on the adjustment dimension was successful.

2. *The Psychotic Group*

The subjects comprising the psychotic group were 30 in number and conformed to the following criteria:

1. Young adults of both sexes with ages from 20 to 40.
2. Persons who were institutionalized and had been diagnosed as paranoid schizophrenics.
3. Superior intellectual potential as evidenced by educational attainment, by prepsychotic vocational status, or by the production of one or more Rorschach responses of high form level.

In all instances there was a history of bizarre behavior including severe departures from reality demands. The severity of such symptoms at the time of testing varied considerably among the 30 subjects. However, with the exception of one patient, none of the subjects was functioning well enough to be under consideration for discharge from the hospital.

Although all members of the psychotic group had been diagnosed as paranoid schizophrenic, many of the subjects presented mixed symptoms and several—because of personality changes since the time of diagnosis—fit more closely into some other entity. Since paranoid subjects were selected solely for the reason that they tend to be less deteriorated and therefore easier to test, it was not necessary to verify diagnoses. It was essential merely that the presence of psychosis be established, and this, of course, was easy to do.

D. THE INSTRUMENTS AND THEIR MANNER OF ADMINISTRATION

Six tests were administered in three different sessions of approximately one to two hours' duration separated by an interval of at least one day. New tests which were devised for the experiment are reproduced in the original dissertation (5). The sequence of testing was as follows:

First session:	Rorschach Digit Test, first administration
Second session:	Information Test Action Picture Test
Third session:	Multiple Choice Stories Test Picture Description Test Digit Test, second administration

1. *The Rorschach*

The Rorschach was used, as described below (II, A), as a basis for securing the criterion of adjustment. It was administered according to the procedures developed by Klopfer and Kelley (4).

2. *The Digit Test*

The purpose of the Digit Test was to obtain a measure of the perception of the self object by requiring subjects to report upon the accuracy of their own performances. The Digit Test made use of the stimulus material of the digits subtest of the Wechsler-Bellevue Adult Intelligence Scale (9, p. 177). Only one trial was given at each level of difficulty instead of the two trials provided by Wechsler. After repeating each digit series, the subject was asked to report upon the accuracy of his own performance. The preliminary instructions were as follows:

> I am going to say some numbers and when I am through I want you to repeat them in the same order as I have said them. After you have repeated the numbers, I shall ask whether you have repeated them correctly or incorrectly. Or perhaps you will be doubtful. You will merely say correct, incorrect, or doubtful.

Before presenting each digit series forward the experimenter said:

> Listen carefully and be sure to say the numbers in the same order as I do.

After the subject had repeated each digit series he was asked:

> Did you say them correctly, incorrectly, or are you doubtful?

Before presenting the digits backwards, the experimenter explained as follows:

> Now I am going to say some numbers and when I am through, I want you to say them backwards. For example, if I say 2, 8, [illegible] you will say 3, 8, 2. As before, I shall ask whether you said the numbers correctly, incorrectly, or whether you are doubtful.

Before presenting each digit series backwards, the experimenter said:

> Listen carefully and be sure to say the numbers backwards.

As before, after each response, the subject was asked to report whether his response was correct, incorrect, or whether he was doubtful.

3. *The Information Test*

The purpose of the Information Test was to provide a situation that disclosed the extent to which subjects were aware of what facts they know and what facts they do not know. In responding to the test items a subject made judgments concerning the conformity of his own beliefs to the standard of objective fact. The test aimed therefore to make manifest, in a limited sense, the manner of perceiving the self object and the congruity of this view of self with a more objective view.

The Information Test consisted of 50 items of the multiple-choice type with four possible choices in each case. In devising the test an effort was made to secure questions from a wide variety of subject fields and to eliminate in this way the possible influence of specialized knowledge and abilities. Questions were chosen which were judged to be familiar to and yet difficult for the average college graduate.

The preliminary instructions to the subject were as follows:

> This is a test to show how much information you know. But it also has another purpose. It shows how aware you are of what you know and what you do not know. Some people believe they know things when they really don't. In taking this test you are supposed to answer only those questions that you are absolutely sure you know. Do not guess. If you are not positive about the answer to a question, leave that question blank. After you are all finished you will have a chance to raise your score by guessing at the questions you left blank. Remember, make as good a score as you can, but don't guess.

The exact wording of the instructions varied somewhat from subject to subject, but the substance was repeated and emphasized until there was no doubt that the subject understood that he was not to guess. After the subject responded to all those items whose answers he was positive he knew, he was furnished with a red pencil and was asked to fill in all the remaining blanks by making the best guesses he could. On the completed test blank all items were marked, either in black or in red.

4. *The Multiple Choice Stories Test*

The Multiple Choice Stories Test consisted of 13 uncompleted stories, each of which was followed by two possible conclusions. The unfinished stories and their conclusions were mimeographed in such a way that the subject's choice of conclusion could be indicated by a check. One of the two conclusions was more probable of occurrence and more consistent with the facts given. Although the test does not directly measure the perception of reality, its inclusion was based upon the hypothesis that one's method of organizing reality is revealed by the dimension of the story to which he responds.

The experimenter is indebted to Wolfenstein for the idea of the Multiple Choice Stories Test (10). Four of the stories used have the same plots as stories of Wolfenstein although the exact wording is not the same. Nine additional stories were composed by the experimenter.

At the beginning of the test administration the subject was handed the test material and permitted to examine it for a few moments. Then he was instructed as follows:

> This test is made up of a number of very short stories. Each story is unfinished, and two possible ways of ending the story are given. You are to read each story and each of the endings which follow it. Then decide which of the two endings you like better. There are no right or wrong answers. Just be guided by your own likes and dislikes.

After a subject had checked all stories, there was an inquiry, the main purpose of which was to determine the basis upon which the choice was made for each story. Responses to the inquiry were recorded verbatim.

5. *The Action Picture Test*

The materials of the Action Picture Test consisted of 11 eight by five inch cards, each card having upon it a simple line drawing of a human figure in a posture of muscular strain or in the midst of an uncompleted act. Those pictures depicting action in the direction of a goal were drawn in such a way that successful achievement of the goal was unlikely.

The subject was furnished with ordinary typing paper and a medium sharp lead pencil. He was given the first card and instructed as follows:

> This is a test to show how accurately you can draw what you observe. Try to copy each picture exactly as it is as nearly as you are able. But do not use a straight edge nor measure the pictures in any way. There will be 11 pictures in all. You may use as much paper as you wish. There is no time limit.

The factors of interest were two: The total amount of discrepancy between the stimulus pictures and the reproductions, and the degree to which the reproduced drawings reduced the muscular tension of the figures or made the successful completion of the act more certain. These tendencies toward reduction of muscular tension and movement toward the goal were called *closure tendencies.* A hypothesis under consideration was that such instances of closure are indicators of a factor of empathetic response. By empathetic response is meant participation in the emotions or actions of another to the extent of actually feeling the emotions and actually contributing one's own motor effort to the efforts of another. Should closure tendencies be indicative of empathetic response, it is logical to assume that they correspond to dissociated and uncontrolled empathy. This follows from the nature of the instructions which direct the subject to copy the figures exactly as they are.

6. *The Picture Description Test*

The Picture Description Test was devised in an effort to examine the degree to which the verbal report of a presented visual stimulus diverges from the objective characteristics of the stimulus. Ten pictures obtained from magazines and mounted on stiff Manila paper 9 × 12 inches in size were used as stimuli. The pictures were highly structured in the sense of representing familiar objects and people in meaningful relationships. Each picture was placed upon a standard and exposed to the subject's view for a three-second interval; after a ten-second waiting period, the subject described the picture as precisely and completely as possible.

The instructions were as follows:

> I am going to show some pictures to you, one at a time. Each picture will be shown for three seconds. After the three seconds are up and the picture is removed, you will have a ten-second period to think about the picture. Then I will ask you to describe it to me. This will show your ability in observation. The more you remember the better your score will be. But be sure to be completely accurate. Do not interpret the picture or add things that are not actually shown. Just describe what you have seen and see as much as you can. Remember, three seconds, then wait ten seconds, then tell me everything you remember. Just before I show a picture, I shall say, "Ready," so that you know it is coming.

The card to be shown was placed on the standard and covered with another card. The experimenter said, "Ready," and removed the covering card for a period of three seconds as measured by a stop watch. This rough procedure resulted in an actual exposure time a fraction of a second more or less than three seconds, but the amount of error introduced is not large enough to require any more precise method of governing the exposure time.

II. SCORING METHODS

A. The Rorschach

The quantitative measure which served as the criterion of the degree of adjustment was derived from Rorschach results, but the criterion used for analysis on the level of the normal-neurotic group was differently obtained than the criterion used for analysis on the level of both groups combined.

1. *The Munroe Check List Rating*

The adjustment criterion used for analysis on the level of the normal-neurotic group was the number of tallies assigned to a Rorschach record on the Munroe revised check list. This is a relatively objective rating system which has been standardized on college students similar in many respects to the members of the normal-neurotic group in the present study. By assigning individuals to one of three levels of adjustment on the basis of the number of check list entries, Munroe obtained a contingency coefficient of .62 between the adjustment ratings and carefully obtained external criteria of adjustment (7).

2. *The Order of Merit Rating*

For analysis on the level of both groups combined an Adjustment Score was obtained by transmuting into units of amount an order of merit ranking of the 60 records on the basis of the efficiency of the subjects' general adjustment. This was used in preference to the Munroe revised check list because Munroe included no psychotics in her standardization study. It was judged to be a feasible procedure because of great range of adjustment represented by the 60 cases. Ranking of the Rorschach records was done by an experienced Rorschach worker who had no access to any data other than the Rorschach protocols, the Klopfer scoring of each record, and the age and sex of each subject. Good adjustment was defined for him as freedom from psychoneurotic and psychotic characteristics and the ability to use one's capacities in a socially effective way.

A rank-order correlation of .91 was obtained between the ranking made by the Rorschach worker and a similar ranking made by the experimenter. The 60 records were also scored by means of the Basic Rorschach Score (BRS) of Buhler, Buhler, and Lefever (1), and a *rho* of .75 was obtained between the BRS and the ratings by the Rorschach worker. The ranking by the Rorschach worker was used in preference to the other two because it was free from the halo effects that might have influenced the experimenter's ranking and it discriminated more effectively than did the BRS between the normal-neurotics and the psychotics.

B. The Digit Test

As with the other tests of the battery, the Digit Test was scored in such a way that high scores corresponded with weak and low scores with strong reality contact. In scoring the Digit Test score points were added under the following conditions:

Digit series repeated accurately but reported as incorrect.

Digit series repeated inaccurately but reported as correct.

Digit series repeated accurately but reported as doubtful.

Digit series repeated inaccurately but reported as doubtful.

Discrepancies between responses and reports existed in varying degrees so that precise scoring called for a system of weighting according to the degree of departure from objective fact. Empirical determination of the proper weightings was impossible under the limitations of the study. However, as with the other tests, a heuristic weighting system based upon a priori considerations and the evidence of the data was adopted. Details of the scoring for the several tests are given in the original dissertation (5).

C. The Information Test

In scoring the Information Test, five different tabulations were made: The number of items marked correctly in black, the number of items marked incorrectly in black, the number of red marks, the number marked correctly in red, and the total correct marks in both red and black. The five different classifications were named and symbolized as follows:

Positive Rights (*PR*)
Positive Wrongs (*PW*)
Guesses (*G*)
Guess Rights (*GR*)
Knows (*K*)

The first three of these scores are self-explanatory. The fourth score (*GR*) is not the absolute number of items marked correctly in red, but is the figure obtained after the usual correction for guessing is applied. The fifth score, *K*, is the sum of *PR* and *GR*. *K* represents the absolute number of items answered correctly without regard to whether the subject is certain or merely guessing.

The *PW's* and the *GR's* were presumed to be instances of reality contact weakness. It is here that the subject was certain of his knowledge when he should have been undecided and undecided when he should have been certain. The sum of *PW* and *GR* could not be used as a measure of reality contact, however, since the absolute size of *PW* and *GR* varies with the size of *K*. For example, individuals with a very high *K* score will have relatively lower *PW* and *GR* scores; the test is so easy for them that relatively few questions raise any doubt in their minds. Likewise, individuals with very low *K* scores will also tend to have very few *PW's* and *GR's*. Many items are completely unfamiliar to them, and such completely unfamiliar items are not ordinarily marked in black nor are they guessed correctly.

How to correct for the size of *K* was the most difficult question in devising the scoring system. The method finally adopted was that which discriminated most significantly between the two groups of subjects. This is expressed by the formula:

$$\frac{PW + GR}{\frac{3}{4}K + g - k}$$

where g is the number of guesses (not to exceed 6), and k is the amount (not to exceed g) by which $\frac{3}{4}K$ exceeds PW. In those records where $\frac{3}{4}K$ was greater than PW plus G, the formula used was:

$$\frac{PW + GR}{PW + G}.$$

Scoring weights for several of the test scores, notably the Information Test score (defined above) and the Deviations Score (described in section F, 2, below) were derived from differences noted between the psychotic group and the normal-neurotic group. The resulting weights were then applied to the same groups from which they were derived instead of to a "cross-validation" sample. An inflation of validity coefficients may have been caused by this procedure, but certain facts suggest that such inflation was minimal:

a. Some of the scores (e.g., the Digit Test score and the Multiple Choice Stories Test score) were not weighted by reference to group differences. Yet these scores have validity equal or superior to scores that might be questioned (cf. Tables 3 and 5).

b. Correlations within the normal-neurotic group would not be greatly affected by the psychotic vs. normal-neurotic group differences that guided the assignment of weights. Yet the *r*'s within the normal-neurotic group are consistently positive (cf. Table 3).

c. Although the Omissions Score (see section F, 1, below), is less directly affected by group differences than the Deviations Score, it is more highly correlated with the criterion (cf. Tables 3 and 5).

Whatever inflation of validity coefficients may have resulted from the weighting methods used, this inflation is probably not a major factor and does not enter into all scores.

D. THE MULTIPLE CHOICE STORIES TEST

Since the choice of story endings in the Multiple Choice Stories Test was governed by different factors in the case of different subjects, a count of the number of realistic endings chosen has no definite meaning. There is no significant difference between the two groups in the absolute number of realistic endings chosen. The real value of the test lies not in any mechanical summation of the choices but rather in its use with inquiry as a method of eliciting the criteria of choice, the subject's particular mode of organizing or perceiving the problem of the story.

All the bases of choice which were reported in the inquiry were listed, and categories for their classification were determined. The resulting categories were then arranged on an a priori basis into four levels of objectivity. The levels of objectivity and the categories they included are as follows:

Level A: (*a*) More probable of occurrence; (*b*) More consistent with the facts given.

Level B: (*a*) Happier or more satisfactory outcome; (*b*) Better for characters' personal development; (*c*) Greater social good; (*d*) In harmony with moral or ethical principles of a universal kind.

Level C: (*a*) Rationalizations (subject argues that the pleasant but unrealistic ending is more plausible, logical, or realistic); (*b*) Personal prejudices, personal likes and dislikes; (*c*) Consistent with dogmatic, provincial, or childish moral principles.

Level D: The reasons for choice are irrelevant to the central problems of the story. This includes absurd rationalizations, impossible cause-effect relationships, and the ignoring or twisting of salient relations of the original story.

A ratings were weighted as zero, B ratings as one, C ratings as two, and D ratings as three.

E. THE ACTION PICTURE TEST

Two scores were obtained for the Action Picture Test: viz., a Distortion Score and a Closure Score.

1. *The Distortion Score*

To obtain the Distortion Score, each drawing was divided into a number of sub-wholes and relationships, and the resulting units were rated piecemeal. Additional points were added when the total gestalt was radically modified.

Each of the component units of a drawing was rated on a five-point scale and assigned values from zero to four. The criteria for assigning the respective values were as follows:

0: No clearly visible alteration.
1: Alteration is clearly visible but not obvious.
2: Alteration is obvious but not marked.
3: Marked alteration.
4: Extreme alteration.

An alteration, to be scored, had to be a clearly visible alteration in relative size, direction, or proportion. No attention was paid to such factors as the amount of pressure, thickness of lines, erasures, or the like. Examples of subject's drawings and their assigned ratings appear in the original dissertation (5).

After rating the component units of a drawing, the drawing was regarded as a whole to judge whether the essential meaning of the stimulus picture was maintained. In some instances, particularly in the records of psychotics, the gestalt disintegrated so that the reproduction was a meaningless collection of parts rather than an organized whole. Two degrees of destruction of the gestalt were quite easily recognized. These were called *partial destruction* and *complete destruction*. In partial destruction some semblance of the original unity remains, but only by effort can the relationships be held together. In complete destruction the gestalt is wholly lost. In cases of partial destruction five points were added to the total rating of a picture; and in the case of complete destruction ten points were added.

2. *The Closure Score*

To ascertain whether closure tendencies were to any extent independent of a general distortion factor, it was necessary to discover whether a subject's tendencies toward closure differed significantly from his tendencies away from closure. The Closure Score was therefore stated in terms of a proportion and was arrived at as follows:

Degrees of error toward and away from closure were rated on a five-point scale according to the same criteria used for rating other alterations in the drawings (II, E, 1). Ratings of errors in the direction of closure were given plus values; ratings of errors in a direction away from closure were given minus values. Instances of partial or complete destruction of the gestalt received a closure rating of minus four. The plus ratings were then divided by the sum of the plus and minus ratings to obtain the Closure Score. In summing the plus and minus ratings, algebraic signs were disregarded. Examples of drawings with their closure ratings appear in the dissertation (5).

F. The Picture Description Test

Initial scanning of subjects' reports to the Picture Description Test disclosed two main differences between the two groups of subjects. The psychotics not only tended to give distorted reports, but their reports in many instances were scantier or omitted details that were reported by the majority of the normal-neurotic group. A scoring system was therefore used that recognized both these dimensions. The two aspects of a record were separately scored, as the Omissions Score and the Deviations Score.

1. *The Omissions Score*

The Omissions Score was based upon the number of frequently reported items omitted from a record. A check list for determining the Omissions Score was prepared, and items on the check list were checked if the corresponding details in the picture were not correctly reported by the subject being rated. Included in the check list were all details that were correctly reported by 50 per cent or more of the normal-neurotic subjects. Items were weighted according to the frequency of their correct report within the normal-neurotic group.

2. *The Deviations Score*

The Deviations Score was a measure of the degree to which a subject's report deviated from the objective characteristics of the stimulus pictures. To arrive at a method for obtaining the Deviations Score, a classification of the various types of deviations was made and the number of each type of deviation was tabulated for each group. Except for certain liberties taken in the interest of consistency, scoring weights were based upon comparative frequencies from one group to the other. For each class of deviation the weighting approximated the ratio of its frequency in the psychotic group to its frequency in the normal-neurotic group. Examples of kinds of deviations and the weighting assigned to them are given in the dissertation (5).

III. RESULTS

A. Reliability of Tests

Table 1 present the split-half reliability coefficients and the standard deviations of the several test scores. Except in the case of the Closure Score, the reliabilities are above .60, and the tests may therefore be considered adequate, in so far as reliability is concerned, for the group comparisons made in the present study. Because of the low reliability of the Closure Score, it was excluded from all further analysis.

For individual comparisons, reliability should be .90 or higher; and for practical usefulness the criterion of .90 should be attained within the normal-neurotic range. The increases in test length required by the Spearman-Brown formula to attain the criterion of .90 within the normal-neurotic range are entirely feasible for all tests except the Digit Test and the Information Test. These tests, if increased to the predicted length, would need to be administered in several sessions in order to exclude the factor of fatigue. In addition to increasing test length, reliabilities might be further raised by the correction of certain weaknesses discussed below (III, B, 1). It may be concluded that, with the exception of the Closure Score, the reliability coefficients are sufficiently high to warrant the further use and development of the tests.

B. Validity of Tests

Consideration of the test validities was approached from two standpoints. Efforts were made first to determine the possible influence of irrelevant variables upon the test scores. Secondly, an attempt was made to ascertain whether the behavior sampled by a test is representative of all or part of what is meant by reality contact in real life situations. In this latter sense a test is valid in so far as it has some meaning apart from the actual task it imposes.

1. *Freedom from Irrelevant Variables*

The problem of eliminating irrelevant variables from the test battery was not equated with a search for primary factors. Even though the tests should be freed from the undue influence of irrelevant factors, the functions measured may still partake of several abilities—and these may turn out to be no different from such classic functions as memory, ability in spatial relations, inductive reasoning, and the like. That this should be true does

TABLE 1

Reliability Coefficients and Standard Deviations of Test Scores in the Normal-Neurotic Group and in Both Groups Combined*

Test	Normal-Neurotic Group		Both Groups Combined	
	r	*SD*	*r*	*SD*
Digit Test	.64	5.6	.83	10.7
Information	.73	29.8	.71	49.1
Multiple Choice	.84	6.3	.89	10.3
Distortion Score	.88	18.2	.93	42.0
Closure Score	.11	20.8	.26	19.2
Omissions Score	.73	8.4	.91	19.7
Deviations Score	.94	22.2	.92	28.8

* (All coefficients have been corrected by the Spearman-Brown formula)

not necessarily imply that reality contact might better be assessed through an isolation and measurement of its constituents. Although reality-testing tasks undoubtedly involve the known mental abilities to a greater or lesser extent, the issue is not whether an individual has these abilities but how well he can utilize them in the service of relating himself to reality. Reality contact is probably not a discrete function in the sense of being independent of other abilities, but it may prove to be a unique function in that its total variance cannot be accounted for by the variance of its known ingredients. The aim in measuring reality contact is to sink a probe into human behavior at an angle different from that of other tests. Though the tests may cut across the same abilities as sampled by other tests, these abilities are sampled in a unique way.

a. The Digit Test. Examination of the Digit Test results indicated that memory span for digits was not excluded as a variable in those subjects who had unusually high memory span for digits. This was because none of the digit series was difficult enough for them to raise any doubt in their minds about the accuracy of their responses. Although this test defect may have resulted in a substantial error of measurement in several cases, the error was necessarily limited to those few subjects who had unusually high memory span for digits. For adequate measurement of such subjects the addition to the test of several longer digit series is necessary.

An important question was whether or not the test score differed appreciably from the memory span for digits. To test this question, a score for the memory span was computed and was correlated with the Digit Test score. An *r* of .51 was obtained. When corrected for attenuation, this was raised to .60. This correlation may be spuriously high because of the test defect mentioned above, but it nevertheless indicates that something in addition to memory span was measured. It is to be expected that reality contact should correlate with memory of all kinds and that this particular test of reality contact should correlate highly with memory span for digits. But the relatively low correlation obtained supports the view that having an ability is not necessarily equivalent to being able to use that ability in relating oneself to reality.

Another problem of interest is whether, when memory span is held constant, any significant relationship remains between the criterion of adjustment and the Digit Test score. Memory span for digits correlated with the Adjustment Score to the extent of .38, while the Digit Test score correlated with the Adjustment Score to the extent of .60. Partialing out memory span resulted in a net correlation between the Adjustment Score and the Digit Test score of .52. It is probable, therefore, that the Digit Test score is not only somewhat independent of memory span but also that it is more relevant to an individual's adjustment than is memory span.

b. The Information Test. The results of an item analysis suggested that eight or ten of the Information Test items introduce a factor of reading comprehension by leading the subject into false premises or by suggesting a fallacious answer. Obviously a subject must understand the question before he is able to make a judgment upon his knowledge concerning the question.

Another problem concerned the extent to which the actual amount of information known varied with the Information

Test score. *K*, which represents the absolute number of items correctly marked, was correlated with the Information Test score and an *r* of .61 was obtained. Corrected for attenuation this was raised to .81. The result indicates that something more than a person's store of information was measured. The rather high correlation obtained is probably somewhat spurious because of test defects. Subjects with good analytic ability and a large fund of information are more likely to understand those questions which are misleading. Because of this understanding they will have a greater number of accurate responses (higher *K*) and a lower Information Test score (indicating higher reality contact). To the extent, therefore, that the Information Test score is contaminated with a factor of reading comprehension the correlation is spuriously high.

The relevance to adjustment of the portion of variance in the Information Test score not accounted for by factual knowledge was also investigated. The *r* between *K* and the Adjustment Score was .39 while the *r* between the Information Test score and the Adjustment Score was .63. Partialing out *K*, a net correlation remained between the Adjustment Score and the Information Test score of .53. There is reason to believe, therefore, not only that factual knowledge and the awareness of one's factual knowledge are nonequivalent, but also that awareness of one's knowledge has more bearing upon general adjustment than does the mere possession of knowledge.

c. *The Multiple Choice Stories Test.* The subject's response to the Multiple Choice Stories Test was not restricted to one dimension, since within the limits of the material he was free to choose the dimension to which he responded. For scoring purposes, however, all responses were ordered to the single dimension of objectivity. It was assumed that, whatever other qualities behavior possesses, it also varies in its degree of objectivity or in the degree of reality contact it manifests. Validation of the scoring procedure requires scaling of responses by a number of judges and other experimentation. However, within the scope of the present study empirical validation of the scale of objectivity appeared in the distribution of responses for the two groups. Of all responses made at the highest level of objectivity, the *A* level, 75 per cent were made by members of the normal-neurotic group. Proceeding down the levels of objectivity, the proportion of responses from the psychotic group increased progressively.

d. *The Distortion Score.* Irrelevant variables that might feasibly influence the Distortion Score are motor skill and the care or effort expended in the task.

It is doubtful, however, whether motor skill in this test situation can be thought of as independent of reality contact. Although subjects can point out discrepancies between their own reproductions and the original drawing, it is not necessarily to be inferred that their errors arise from motor disability rather than from reality contact weakness.

> Under the lenient scoring system used, a person with normal motor control can reproduce the drawings accurately if he perceives them accurately. The reason individuals make errors which they can later detect as errors is probably because visual perceiving (or the mere contemplation of an object) does not result in the same perceptual emergent as does the visual motor effort of drawing the object. The strong involvement of the motor functions makes for a different sampling of the organism's energy. With some subjects the greater participation of the motor apparatus tends to increase the influence of the inner personal organization causing a greater deviation from the stimulus

object. In other individuals the opposite may be true: Involvement of the motor functions produces a product more in harmony with the stimulus object.

Although capacity for motor control probably has only a negligible effect upon the Distortion Score, a factor of effort is undoubtedly involved. By tentative sketching and frequent erasures certain subjects were able to produce a relatively flawless copy of the original. They improved their visual-motor product by checking and rechecking, perhaps by comparing it with a purely visual perception. With tachistoscopic projection and drawing from memory this factor could be more adequately controlled.

To make the most efficient use of visual motor tests, it is necessary to identify their principal qualitative dimensions. Although distortions may be measured in terms of degree, they also vary in kind. Because of the various qualitative possibilities for response, the total distortion score is a rather crude measure. Nevertheless it is a meaningful measure in that it varies in a general way with adjustment.

e. The Closure Score. The dimensions represented in the Closure Score can only be guessed at since the score itself lacked a sufficient measure of inner consistency. It is probable that the Closure Score is the effect of multiple causes and that these causes operate in varying degrees with different pictures and with different subjects. Negative closure errors, for example, might indicate a bending over backward to check impulsiveness or an actual unresponsiveness to the relationships of the drawing. Another complication is that closure involving the human figures may compete with other closure possibilities of a picture so that what is a good figure from a design sense may not be a good figure from the sense of goal-directedness.

f. The Picture Description Test. From the standpoint of the intrusion of irrelevant factors, there seems to be no serious defect in the Picture Description Test. The Omissions Score, which is a measure of frequently seen items not correctly reported, is qualitatively different from the other measures used in the study. The other scores are chiefly concerned with the degree to which objective events are subjectively distorted. The Omissions Score is a measure of the degree to which one is aware of those objective relationships which are perceived by most people. Obviously, since the scoring system for computing the Omissions Score is based upon deviations from normative standards, the validity of the measure depends upon a scoring system developed from a much more extensive sampling than that of the present study.

2. *Wider Applicability*

Are the measurements obtained specific to the conditions of measurement or do they have predictive value for other situations? Findings relevant to this question are reported in Table 2. Additional findings, which concern the relationships between test scores and the criterion of adjustment, are reported later (III, C).

Table 2 reports the test means and standard deviations for each group, together with the critical ratios of the differences between the means on each test. It also reports for each test the percentage of the normal-neurotic scores that overlap the median of the psychotic group and the percentage of psychotics' scores that overlap the median of the normal-neurotic group. In addition to these data for individual tests, the same data are given for a Composite Score. The Composite Score represents the average of the T scores of the several tests of reality contact (excluding the Closure Score). The last column of the table reports the correlation of each test with the Composite Score.

The findings reported in Table 2 may be summarized as follows:

1. The hypothesis that there is no

TABLE 2

MEAN AND *SD* OF EACH TEST SCORE AND OF THE COMPOSITE SCORE, CRITICAL RATIOS, PERCENTAGE OVERLAP, AND *r* OF EACH TEST WITH THE COMPOSITE SCORE

Score	N-N. Group		Psychotic		CR	% N-N. o'laps Psy.	% Psy. o'laps N-N.	*r* with Compos. Score
	Mean	*SD*	Mean	*SD*				
Digit	9.87	5.58	21.2	11.46	4.80	10	13	.71
Information	76.8	28.6	135.6	61.02	4.71	7	7	.73
Omissions Score	23.5	8.43	52.0	11.94	10.48	0	0	.82
Deviations Score	26.33	22.2	42.0	32.52	2.16	13	33	.54
Distortion	53.0	18.24	99.34	32.6	6.69	0	13	.79
Closure	48.53	20.82	43.0	17.06	1.14	47	27	
Multiple Choice	12.67	6.30	26.8	8.22	7.42	3	10	.87
Composite	44.87	2.66	55.17	6.16	8.31	0	0	

significant difference between the two groups is rejected beyond the 1 per cent confidence level for all test scores except the Picture Description Deviations Score and the Closure Score. The hypothesis is rejected at about the 5 per cent level for the Deviations Score. The difference between the means of the Closure Score is not significantly greater than chance.

2. There is a tendency for scores of the psychotics to overlap the median of the normal-neurotics to a greater extent than normal-neurotic scores overlap the median of the psychotics.

3. The Composite Score shows less overlapping in either direction than the average overlapping of the scores that compose it.

4. All of the tests comprised in the Composite Score correlate with the Composite Score to an extent significantly greater than that due merely to the part-whole relationship involved.

The first finding—the rejection of the hypothesis of no difference between the means of the two groups—supports the hypothesis of wider applicability. The most plausible inference is that the differences noted are due to a relationship between test scores and degree of psychopathology. Any other variables that might logically be related to the test scores are substantially equal in the two groups.

The finding that the scores of the psychotics tend to overlap the scores of the normal-neurotics to a greater extent than the reverse is true suggests that a high score on a single test is a stronger indicator of pathology than a low score is of the lack of pathology. The overlapping of the two groups may be partially accounted for by errors of measurement, but several of the overlapping scores deviate too greatly to be explained by error of measurement. This fact is not disconcerting since it is to be expected that some psychotics will display greater reality contact at specific times and with respect to specific situations than will certain neurotic individuals. The relatively large overlap and poor differentiating power of the Deviations Score is cancelled out if the Omissions and Deviations Scores are combined. Certain psychotic subjects were able to attain a low Deviations Score by withdrawing from the situation, but in so doing they accumulated a high Omissions Score.

The significant correlations found between individual tests and the Composite Score also support the hypothesis of wider applicability. It is noted that the tests which correlate most highly with the

Composite Score—the Multiple Choice Stories Test, the Omissions Score, and the Distortion Score—are also highest in the other criteria of validity. Their reliabilities are high and they show the greatest differentiating power in terms of critical ratios and degree of overlap.

C. Correspondence of Test Scores and Adjustment

1. *The Normal-Neurotic Group*

The results pertaining to the relationship between test scores and the Munroe check list ratings within the normal-neurotic range are presented in Tables 3 and 4. Table 3 reports the intertest correlations within this range. Table 4 presents the successive multiple R̄'s obtained after the addition of each test, according to the Wherry-Doolittle method of test selection. Since, as is seen from the table, the addition of a fifth test lowered the correlation, there was no point in carrying the computation any further.

Conclusions derived from the results reported in the two tables are as follows:

1. All test scores correlated positively with the criterion, though only the Omissions Score showed a statistically significant relationship. The r between the Omissions Score and the criterion is beyond the 1 per cent level of significance.

2. The intercorrelations of the several tests were for the most part nonsignificant, indicating a high degree of test specificity within this range of adjustment.

TABLE 4

Successive Multiple $\bar{R}$'s with the Adjustment Criterion Resulting from the Addition of Tests by the Wherry-Doolittle Test Selection Method (Normal-Neurotic Group)

Tests Added	Resulting $\bar{R}$
Omissions Score	.57
Information Test	.64
Multiple Choice	.67
Deviations Score	.69
Digit Test	.68

3. The r of .57 between the Omissions Score and the criterion, as well as the multiple R̄ of .69, support the hypothesis that reality contact is related to adjustment within the normal-neurotic range. Additional evidences of a relationship are reported below (III, C, 1, *a*).

a. Relation of tests to the criterion It is premature to draw conclusions from such a small group and without a more carefully obtained criterion. However, comparing the multiple R̄ of .69 with Munroe's contingency coefficient of .62 between her check list and carefully obtained validating material (7), it would seem very worth while to develop the tests further. The multiple R̄ was obtained with only three tests. Although

TABLE 3

Intercorrelations of Test Scores and the Munroe Adjustment Rating in the Normal-Neurotic Group

Scores	Digit	Omissions	Deviations	Information	Distortion	Multiple Choice	Adjustment
Digit		.067	−.088	.283	−.033	.588	.338
Omissions Score	.067		.295	−.173	.211	−.141	.571
Deviations Score	−.088	.295		−.041	−.076	−.235	.297
Information	.283	−.173	−.041		.084	.261	.219
Distortion	−.033	.211	−.076	.084		.276	.292
Multiple Choice	.588	−.141	−.235	.261	.276		.228
Adjustment	.338	.571	.297	.219	.298	.228	

four scores were used, the Omissions and Deviations Scores are both obtained from the Picture Description Test. The three tests which produce the multiple $\overline{R}$ are all relatively objective and are adaptable to group administration.

The low correlations between the criterion and all measures except the Omissions Score may mean in part that the functions measured by the tests have very little relationship to adjustment within the normal-neurotic range—although, as already shown, they do have power to differentiate between psychotics and nonpsychotics. It is probable that some of the dimensions contained in the tests are related in an inverse or curvilinear way to adjustment within the normal-neurotic range. Curvilinearity could not be established, but analysis of individual records made evident that certain subjects achieved strict adherence to objectivity by special effort. They clung to reality with a tenacity resembling a neurotic defense mechanism. However, when real perceptual disturbances are present, the drive for objectivity does not actually produce objectivity.

Although test scores were frequently relatively lower (stronger reality contact) than the adjustment rating, scores conspicuously higher than the adjustment rating rarely occurred. This was demonstrated statistically in the following way:

The highest test score of each subject was correlated with the adjustment criterion. The resulting r was .58.

The lowest test score of each subject was correlated with the adjustment criterion. The resulting r was .25.

This evidence suggests that low scores are less indicative of good adjustment than high scores are symptomatic of maladjustment. The actual test score that is high does not matter. A real weakness in reality contact at *any* point is indicative of poor adjustment. It must be borne in mind, however, that the difference between the coefficients of .25 and .58 is significant at only about the 10 per cent level; for verification a larger sample is needed. Nevertheless, the r of .58 is highly significant and is further evidence of a real relationship between reality contact, as measured, and adjustment within the normal-neurotic range.

There was also further evidence of the relation of test scores to adjustment. When separate tests were combined into a composite score, the correlation with the criterion was increased. Thus, the average of the T scores of five tests (excluding the Omissions Score and the Closure Score) gave the significant r of .53 with the criterion, even though none of the scores contained in this average had a greater r than .34 with the criterion. To predict adjustment within the normal-neurotic range, it appears necessary either to combine a number of specific tests or to construct tests that sample a larger area of the phenomenal field. The success of the Omissions Score is probably partially due to the fact that it presents a more complex reality-testing task.

b. The Omissions Score. Since the Omissions Score is related to the criterion in a significant way, the nature of this measure is important to understand. The Omissions Score represents the number of frequently reported items in the Picture Description Test that are not correctly reported by a subject. It is sensitive both to a narrowing of the perceptual field and to tendencies to organize the perceptual field in unique ways. Since the Omissions Score correlates very highly with the total number of items correctly reported (r equals .87), the main factor measured is probably the degree of constriction of the perceptual field. The theoretical significance of this finding is in its consistency with those theories which suggest that when the organism's capacities are reduced by injury, disease, or senescence, it maintains equilibrium by return to a lesser level of organiza-

TABLE 5
INTERTEST CORRELATIONS IN THE COMBINED GROUP

Scores	Digit	Omissions	Deviations	Information	Distortion	Multiple Choice	Composite	Adjustment
Digit		.55	.08	.54	.44	.54	.71	.60
Omissions Score	.55		.26	.53	.63	.59	.82	.66
Deviations Score	.08	.26		.13	.35	.47	.54	.45
Information	.54	.53	.13		.36	.55	.73	.63
Distortion	.44	.63	.35	.36		.61	.79	.55
Multiple Choice	.54	.59	.47	.55	.61		.87	.73
Composite	.71	.82	.54	.73	.79	.87		.80
Adjustment	.60	.66	.45	.63	.55	.73	.80	

tion. Because the organism cannot cope with intruding stimuli, it withdraws or incapsulates and only in this way maintains equilibrium. The process has been referred to by Goldstein as the "catastrophic reaction" (2, pp. 117, 136, 318), by Snygg and Combs as "tunnel vision" (8, p. 125), by Lewin as "encysting" (6, p. 94).

2. *Both Groups Combined*

Table 5 presents the intercorrelations of the several test scores, the Composite Score, and the Adjustment Score in the combined (normal-neurotic and psychotic) group. Comparison of these data with the analogous data of the normal-neurotic group (Table 3) discloses tendencies concerning test specificity and generality:

a. The higher order of interest correlations in the combined group suggests a greater degree of generality for the tests when the analysis covers the total range of adjustment.

b. The higher order of correlations between tests and criterion in the combined group suggests that adjustment is dependent in a broad way upon reality contact even though minor differences in reality contact may not be directly reflected in adjustment.

c. The greater correspondence of the adjustment criterion with the Composite Score than with any of the individual test scores supports the finding in the normal-neurotic group that the more generalized the measure of reality contact the stronger is its relation to adjustment.

Although the significance of these tendencies is doubtful, particularly since two unlike groups are combined, the conclusion may be tentatively drawn that test scores operate in a fairly specific way within a limited range of adjustment, but that an individual's level of adjustment sets certain broad limits above and below which test scores cannot go. The implications of this conclusion are discussed more fully below (IV, A, 1).

IV. IMPLICATIONS OF FINDINGS

A. Theoretical Significance

1. *General Discussion*

The evidence for intra-individual variability in test scores and therefore a certain amount of specificity of reality testing functions is quite clear. Nevertheless this intra-individual variability is relatively small when compared with the range of test scores in the total sample. The degree to which the scores of the psychotics overlapped the scores of the normal-neurotic group and vice versa was relatively small. It may be tentatively concluded, therefore, that some general factor present in the tests of the present battery had a limiting effect upon the range of scores produced by a single individual. The finding of limited intra-individual variability is in line with the idea that specific kinds of reality contact are differentiated parts of the phenomenal field and as such are derived from and dependent upon the total state of the phenomenal field. The argument that these results can be explained just as easily by assuming common elements in the several tests cannot be disposed of, however, without further evidence.

In responding to the tests of the present battery (excepting the Multiple Choice Stories Test) the subject presumably attempts to adapt his own subjective experience as closely as possible to the requirements of objective reality. On highly specific tests a subject's phenomenal experience may be much closer or much farther from objective reality than his average performance, but on tests that are more general in nature, his performance will naturally be nearer to his own average level. To measure a subject's "general" or "average" reality contact weakness, the mean of a number of specific reality contact measures may be taken, or tasks can be contrived which sample a larger area of the phenomenal field. The Picture Description Test, which involves the perception of people and familiar objects in meaningful relationships, probably samples a large area of the phenomenal field. The same is true of the Multiple Choice Stories Test which calls for a consideration of real-life problem situations.

Testing reality contact weakness in specific situations may be conceived of as measuring the distance between the point representing objective reality and the point on the reality dimension at which the task is actually performed. We assume that the subject tries to order his experience as close to the point of objective reality as he can. The validity of this assumption is somewhat questionable, for although the subject is instructed to be as accurate as possible, the task is not crucial to his existence, and he undoubtedly permits a measure of laxness. Some subjects, however, apparently do regard the tasks as crucial to their existence (an unrealistic attitude) and attack them with deadly seriousness. This difference in attitude is usually revealed by internal evidence. It has the greatest effect upon the Distortion Score and the least effect upon the Omissions Score.

With the Multiple Choice Stories Test it cannot be assumed that the subject is even trying to order his experience to the demands of objective reality. In this test the subject is not told to be accurate or realistic but is permitted to seek his own most comfortable point on the reality dimension. The test is therefore useful in determining how close to objective reality the subject orders his experience when left to his own devices. A strict adherence to objective reality under these conditions is not a favorable sign but appears to be the response of a cramped, prosaic, and overconforming individual. Under these conditions of free choice, however, there is a strong tendency for the level of reality chosen to bear a consistent relation to the person's "general" or "average" reality contact weakness (r between the Multiple Choice Stories Test and the Composite Score equals .87).

2. *Reality Contact in Neurosis*

The frequent assumption that there is no difference in perceptual ability between normals and psychoneurotics is not upheld by the results of the investigation. Four different kinds of evidence point to a relationship between perceptual contact with reality and the degree of adjustment within the normal-neurotic range. These include: (*a*) the multiple $\overline{R}$; (*b*) the r with the Omissions Score; (*c*) the r with the highest score of each subject; (*d*) the r with the average of five scores which by themselves showed no significant relationship with the criterion (III, C, 1).

The general inability in the past to establish the presence of perceptual disturbances among psychoneurotics has probably been due either to the use of methods that are not sensitive enough to reveal the disturbances or to the use of measures that are too specific. Differences between normals and psychoneurotics in the perceptual contact with reality are usually too slight to be evident under ordinary conditions of observation. The stimulus must be reduced to a minimum by tachistoscopic methods or the task must be made difficult in some other way. It is evident also that a large area of the phenomenal field must be sampled. Although individual, specific measures show very little relation to adjustment within the normal-neurotic range, when such specific measures are combined into a composite score a significant relationship appears. The positive results obtained with the Omissions Score are probably due both to the limited exposure method used and to the fact that it samples a larger area of the phenomenal field.

B. Applicational Significance

The idea that reality contact can be regarded as a quantitative feature which cuts across all qualitative dimensions of behavior is upheld to a large extent by the findings of the study. Reality contact undoubtedly fluctuates under differing conditions, but the data of the study nevertheless suggest that a concept of general reality contact, even though it be abstracted from particular instances, can probably furnish a conceptual tool much needed for an appraisal of the extent of maladjustment. Although the investigation of specific kinds of reality contact should not be neglected, the more generalized measures of the present study were shown to have the greatest predictive value. Both in the normal-neurotic group and in the combined group, the composite measures showed the greatest correspondence with the criterion of adjustment. Also, those tests which correlated most highly with the Composite Score (the Multiple Choice Stories Test and the Omissions Score) were most strongly related to the criterion of adjustment.

The adjustment criterion, since it was dependent upon a variety of Rorschach factors, should be expected to correlate most highly with other generalized measures. Nevertheless, obtaining significant correlations between a general reality-contact score and a general Rorschach adjustment score implies a validity for both measures. If general adjustment were not a valid concept to begin with, the validity of a general reality-contact score could not be demonstrated. And although specific areas of maladjustment must be located in order to apply psychotherapeutic correctives, the practical value of a general estimate of adjustment efficiency is not to be denied.

The battery in its present form is adequate for distinguishing psychotics from neurotics in all but a few borderline cases, but it is not sufficiently sensitive to neurotic disturbances to make valid conclusions regarding a particular individual. A rather maladjusted person will occasionally produce a low Composite Score (indicating strong reality contact) so that, if the battery were to be used for screening purposes, certain malfunctioning individuals would go undetected. However, in the present data, a conspicuously high Composite Score always indicates severe disturbance of some sort.

Reality-contact measurement may supplement but is unlikely to supplant the usual diagnostic techniques. Practical advantages include the objective nature of the tests and their adaptability to group use. Should the possibilities suggested by the present data be fulfilled in later experimentation, a way to gauge the extent of maladjustment will be established. Yet —whether or not the relationship between maladjustment and reality contact is universally valid—psychodiagnosis obviously implies more than a determination of the degree of malfunction. Also, it is wholly probable that in certain maladjustments—perhaps in the affective disorders—real pathology may exist without significant reality-contact loss. On the other hand, perceptual deviations may in such cases be merely more specific or less observable.

Although not reported in this paper, the data of the present investigation also suggest that specific kinds of reality-contact weakness are associated with specific behavioral tendencies. Several hypotheses along this line are offered in the original dissertation (5). Knowledge of whether reality-contact measures can function in precise personality description and in differential diagnosis depends upon the testing of such hypotheses. And this endeavor depends in turn upon well-developed, specific tests and criteria, which should, perhaps, be closer to actual-life reality than are the laboratory-type exercises of the present battery.

V. SUMMARY

FIVE tests were devised for the purpose of measuring weakness of the reality contact and were administered to a group of 30 paranoid schizophrenics and to a heterogeneous group of 30 subjects, both normal and psychoneurotic. Seven scores were derived from these five tests. The Rorschach was also administered and was used as a basis for estimating the general level of adjustment of each subject. Questions to be investigated were the reliability and validity of the several tests of reality contact, and the relation of the measures of reality contact to the level of adjustment.

A. Results

1. The split-half reliability after correction by the Spearman-Brown formula was sufficiently high to justify use of all scores, except the Closure Score of the Action Picture Test, for group comparison. The Closure Score was therefore omitted from all further analysis. The reliability coefficients of the remaining tests varied from .64 to .94 in the normal-neurotic group, and from .71 to .93 in both groups combined (normal-neurotic plus psychotic).

2. With respect to the freedom of the tests from irrelevant factors, data were insufficient to give a clear-cut answer. It was shown, however, that the Digit Test measured something in addition to memory span for digits, and that the Information Test measured something in addition to the actual amount of information known. An irrelevant variable that appeared to influence the score on the Information Test was the factor of reading comprehension.

3. Significant differences between the mean of the normal-neurotic group and the mean of the psychotic group were obtained in all scores except the Closure Score. The critical ratios of these significant differences varied from 2.16 to 10.48. Several of these figures may have been inflated, though probably not appreciably, by the method of weighting scores (II, C).

4. Significant relationships were found between the measures of reality contact and the degree of adjustment within the normal-neurotic range. The *r* between the Omissions Score and the criterion of adjustment equaled .57. A multiple $\overline{R}$ representing the correlation between the criterion and three tests (four scores) equaled .69. The *r* between the highest score of each subject and the criterion was .58. The *r* between the average of five scores (excluding the Omissions Score) and the criterion was .53, even though none of the scores contained in this average correlated with the criterion to an extent greater than .34.

5. Although there was a large degree of generality or correlation among the several tests in the full-range sample (both groups combined), in the normal-neurotic group the tests were much more specific (uncorrelated) in their operation.

B. Conclusions

1. The experimental results support the belief that relatively objective and reliable tests of reality contact can be devised.

2. The results of the study support the view that such tests possess a validity that extends beyond the specific conditions of measurement.

3. The data of the study uphold the idea that the weakness of the reality contact is related to the general level of maladjustment, both within the normal-

neurotic range and in the total range of adjustment.

4. Within the limits of the study, reality contact appeared to be both specific and general in its operation. Although a certain amount of intra-individual variability existed from one reality testing task to another, the intra-individual variability was relatively small when compared to the range of test scores within the total sample.

5. The evidence of the study supports the idea that a concept of general reality contact has practical usefulness even though it be abstracted from specific instances.

6. The generally positive results of the experiment favor the further refinement of the instruments used, the development of new instruments, and the further investigation of the dimensions and correlates of reality contact.

REFERENCES

1. Buhler, C., Buhler, K., & Lefever, D. W. *Development of the basic Rorschach score with manual of directions.* Los Angeles: Mimeographed by authors, 1949.
2. Goldstein, K. *The organism. A holistic approach to biology derived from pathological data in man.* New York: American Book Co., 1939.
3. Helson, H. The psychology of gestalt. *Amer. J. Psychol.,* 1925, 36, 342-370.
4. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* New York: World Book Co., 1942.
5. Kreinheder, A. *The strength of the reality contact, an attempt at its objective measurement.* Unpublished doctoral dissertation, Claremont Graduate School, 1951.
6. Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
7. Munroe, R. L. *Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method.* Stanford, Calif.: Stanford Univer. Press, 1945.
8. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949.
9. Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.
10. Wolfenstein, M. The reality principles in story preferences of neurotics and psychotics. *Charact. & Pers.,* 1944, 13, 135-151.

Vol. 66, No. 12 Whole No. 344, 1952

Psychological Monographs: General and Applied

The Performance of Schizophrenic and Normal Individuals Following Frustration*

Harold Wilensky

Franklin D. Roosevelt Veterans Administration Hospital, Montrose, New York

I. PROBLEM AND METHOD

Frustrating situations have been stressed by many authorities as a major area in which the reactions of schizophrenic and other psychotic persons differ from the normal. The theories relating frustration and schizophrenia, however, appear to be based mainly upon clinical impressions, since the studies reported in the literature offer little evidence to support or detract from the various hypotheses. (For a bibliography, see reference 15.)

The present study was initiated to provide more nearly adequate experimental evidence regarding the difference between the frustration tolerance of schizophrenic and normal individuals. The major hypothesis in the present study is that the schizophrenic group will manifest a lower frustration tolerance (i.e., a greater quantitative reduction in their level of performance) than the normal group following failure to attain success within the specified tasks. A second hypothesis is that the responses of the schizophrenics will differ qualitatively from those of the normals during and following frustration—for example, they will give a greater number of bizarre responses and will be more apt to abandon the task than the normals. A third hypothesis is that as the number of tasks in which failure is encountered increases, the cumulative effect of these successive frustrations will result in a progressively greater reduction in performance, and this cumulative effect will be greater in the schizophrenic group than in the normal group.

For the purposes of the present experiment, frustration was said to exist when (*a*) the individual had demonstrated a need to attain success in a task, and (*b*) success in certain subsequent items of the task was thwarted by virtue of the difficulty of the objective situation. The subjects were merely informed of failures when they occurred. In order to infer the existence of at least a minimal need for success (making the correct response) the subject had to perform adequately on three items in the initial form of each task (i.e., obtain a minimal prefrustration score equivalent to three correct responses).

---

* An abridgement of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Graduate School of Arts and Science, New York University.

The author wishes to express his sincere appreciation to Professor Leland W. Crafts for his guidance throughout the course of this study.

He is also grateful to Dr. Earl E. Swartzlander, former Chief Clinical Psychologist, and to Dr. Louis Verdell, Manager, both of the Northport Veterans Administration Hospital for their permission and encouragement to carry through this experiment, and to Dr. Robert S. Morrow, Chief Clinical Psychologist of the Bronx Veterans Administration Hospital, for his assistance in completing the study.

For their helpful criticisms and their discussion of the many problems encountered during the course of the research, acknowledgement is due to the many colleagues of the author in the Veterans Administration.

For her patience and encouragement, the author wishes to express his deep gratitude to his wife.

The study was reviewed in the Veterans Administration and is being published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

A. Materials

Four tasks, each consisting of two equivalent forms, were employed. Items within each task were graded for difficulty so that most individuals were likely to experience success initially and encounter failure on the later more difficult items. The tasks as administered to the main experimental groups are described below.

1. Auditory Memory. This task (Digit Span) consisted of sets of digits of increasing length. The subjects were instructed as follows: "I am going to say some numbers. Listen carefully, and when I am through, say them right after me."

The experimenter read the digits at the rate of approximately one per second, beginning with the shortest series, which consisted of two numbers. If a subject repeated the digits correctly, the next longer series was given without comment by the experimenter. In the case of incorrect repetition the subject was permitted three trials for a given series length before it was considered failed. The subjects in the main experimental groups were informed of failure whenever it occurred by the statement: "That was wrong; now listen carefully and be sure to get them just right."

Series up to a length of two digits beyond their just-determined memory span were administered to the subjects, providing a minimum of six successive failures. Following these failures the alternate form was administered in descending order of difficulty beginning with a series length one digit beyond the just-determined memory span to allow for possible improvement. The next shorter series was given when a subject had failed a set of three trials or had correctly repeated one trial. The descending order was terminated when the shortest series had been reached.

Three points were given for the correct repetition of any given length of digits on the first trial; two points for the correct repetition on the second trial; and one point for the correct repetition on the last trial. The score obtained on the form of each task administered in ascending order of difficulty (until the specified number of failures was encountered) is hereafter referred to as the *prefrustration* score; the score obtained on the form of each task following failure is referred to as the *postfrustration* score. The *drop* score (postfrustration score subtracted from the prefrustration score) was calculated for each subject. In each of the following tasks drop scores also were calculated in an identical manner.

2. Visual Memory. The Benton Visual Retention Test (1) was employed. It consists of two forms of seven cards each, upon which one or more geometrical figures are drawn in india ink. The subjects were instructed as follows: "I am going to show you some cards on which there are one or more designs. You will be given a short time to study each card. Then it will be removed and you will draw what you have seen." The subject was given a pencil and blank paper approximately the same size as the cards (5 × 8 inches). At card number three (the first card with more than one design) the subject was reminded: "Don't forget to draw everything you see."

Each card was presented to the subject for five seconds and then removed. (Standard administration time is ten seconds.) Because of the brevity of this task and the likelihood of partial successful reproduction even on the difficult items, all seven cards in each form were administered. The prefrustration form was administered in ascending order of difficulty and the postfrustration form in descending order. When the subject incorrectly reproduced or omitted any part of the figures on a card, he was told: "That was wrong; look at the next design carefully and be sure to draw everything you see." In the case of correct reproduction no comments were made.

A more sensitive scoring system than the one used by Benton was developed which took into account the shape of the figures (or figure), their horizontal order, their vertical position, and their relative size. Approximately two points were scored for each figure, less points subtracted for inaccuracies.

3. Auditory Perception. The test consisted of two recordings of 54 spondaic words (a term used in verse referring to a foot of two long syllables) at nine graded intensity levels. The words were selected from a list of 70 spondees prepared by the Harvard Psycho-Acoustic Laboratory (16). The list originally was intended to supply a relatively large number of homogeneous items to measure the threshold of hearing for words. Each form of the test designed for the present study consisted of 27 different spondees recorded at nine different levels of intensity. There were three words in each of the nine different steps.

In the recording process all of the 54 words were spoken at a constant level through a calibrated amplifier into the recording device. After each step of three words was spoken, the intensity level of the amplifier was reduced by three decibels until the nine steps (27 words) in the first form had been recorded. For the second form the intensity level of the amplifier was increased after each step by three decibels. Another recording in which the forms were alternated was made in an identical manner to permit rotation of the forms.[1]

A portable phonograph equipped with two sets of headphones was used to play the recording during the experiment. The headphones worn by the subject were of the type used by the Army Air Forces. They had two large sponge rubber cups which fitted over the ear pieces and tended to reduce external noises. The experimenter wore a single earphone to monitor the recording. The subjects were instructed as follows: "I am going to play a recording of some spoken words. Your job is to repeat the word as soon as you hear it. If you have any difficulty, you should guess because you may get it right. They are all familiar, meaningful words." The subject then was given the headphones.

There were intervals of approximately three seconds between the words to permit repetition of the word by the subject. Usually, an interval of six seconds was allowed between successive steps, but when comments by the experimenter were required, the interval was longer.

The volume control was set initially at a low intensity level (a point determined during trial runs at which none of the persons tested could identify more than half of the six syllables (three words) in the seventh step of the first form. Those subjects who correctly identified the three words in the first step continued on to the next steps; the volume-control dial remained fixed. When a subject encountered difficulty in identifying all three words in the first step, the same step was repeated with the comment: "Let's try it again to get used to it." If a subject still was not successful, the volume was gradually increased until a point was reached where he could identify all the words in the first step. The test then was continued and the volume-control dial remained fixed at the new position throughout the presentation of the two forms.

When a subject had incorrectly identified or omitted three of the six syllables within a step (after the first), the turntable was stopped and the subject informed of failure. In the case of incorrect responses the subject was told: "That was wrong; now listen carefully and be sure to get the next ones right." If no responses were made, the subject was asked: "What words did you hear?" If incorrect responses were given, he was informed of his errors. When no guesses were offered, he was told: "You missed those words; now listen carefully and be sure to get the next ones right." No comments were made when more than three syllables in any one step were identified correctly. The recording was played in its entirety for all subjects in the main experimental groups regardless of the subjects' limens.

One point was given for each syllable correctly identified. The first step of the prefrustration form and its equivalent step in the postfrustration form (i.e., the loudest step) were given full credit since all subjects were required to respond correctly to the very first step.

4. Visual Perception. The test consisted of two forms of 15 blurred words arranged in order of difficulty of identification. In order to control for differences in the familiarity of words, the thirty words were selected from the thousand most frequently used words in the Thorndike and Lorge lists (14). Each word was seven letters long and no letter was repeated within any one word. Varying grades of blurredness were obtained by typing the words on bond paper and making 14 carbon copies. Six different words were typed on each sheet; they were typed in capitals with the letters an additional space apart to prevent them from merging into each other in the blurring process. The sixth, eighth, tenth, twelfth, and fourteenth copies were used. The five carbon copies (30 different words) then were photographed on 35-mm. negatives. In the reproduction process the words were enlarged to approximately original size. Each word was on a separate 3×4-inch card of doubleweight, glossy enlarging paper. The order of the difficulty of identification of the words was determined by the performance of a group of normal subjects. The words were placed in rank order and divided into two equivalent forms.

The subjects were told: "I am going to show you some photographs of words. These words will be blurred, but your job is to read each word. If you have any difficulty in reading these words, you should guess because you may get them right. They are all familiar, meaningful words. For example, read this one." The subject was given the first card. Whenever a subject indicated that he could not respond, he was instructed: "See what letters you can make out and then take a guess." No comment was made if a subject correctly named four of the seven letters or gave a seven-letter word with four or more letters correctly identified in the proper positions. If a word other than one with seven letters was given, the subject was asked to spell the word. Usually the subject indicated which letters he had used.

---

[1] The author is grateful to Dr. M. Bergman, Assistant Chief of the Audiology Clinic at the Veterans Administration New York Regional Office, for making the recordings.

When a subject incorrectly identified four or more letters in a word, he was informed: "That was wrong." (In the case of no response: "You missed that one.") "Be sure to study the next word carefully." After the first failure, the correct word was named by the examiner. Since a blurred word can generally be recognized after it has been identified, this procedure tended to assure most subjects that the stimuli consisted of real letters and were not merely "smudges," or "fingerprints," or "ink blots."

The main experimental subjects were allowed three successive failures in the prefrustration form (ascending order of difficulty) before being given the postfrustration form in descending order of difficulty. The postfrustration form began with a word at the same level of difficulty as the last word failed in the prefrustration form. A score of one point was given for each letter correctly identified.

B. Preliminary Experiment

1. Purpose. A defect in one of the previous frustration studies (2) was the lack of consideration given to initial differences in behavior between schizophrenic and normal individuals. For example, even in nonfrustrating situations it has been shown that schizophrenic individuals are unable to maintain effective mental sets for as long a period as normals (12). In addition schizophrenic patients are more likely to discontinue tasks of their own accord (10). Attention and concentration difficulties also are characteristics of schizophrenic behavior. Despite the brevity of the experimental tasks, these behavior tendencies might result in a decrement in performance merely due to the length of the tasks. It was necessary, therefore, to determine the extent to which such variables related to the length of the tasks might affect adversely the performance of schizophrenic subjects under conditions of minimal frustration.

2. Population. The 24 male patients employed in this preliminary study were obtained from the population diagnosed as schizophrenic at the Veterans Administration Hospital, Northport, New York. They ranged in age from 22 to 35, with a mean age of 27.6 years. The patients had been hospitalized from several months to over seven years, with a mean length of hospitalization of 3.5 years. All subjects had been raised in New York or immediately neighboring states and had completed from 8 to 15 grades in school, with a median educational level of 11.0 grades. Their IQ's, roughly estimated by prorating their combined Vocabulary and Information subtest scores of the Wechsler-Bellevue Scale, ranged from 92 (a combined weighted score of 17) to 127 (a combined weighted score of 27), with a mean of 110 (a combined weighted score of 21.9).

3. Procedure. After administration of the Vocabulary and Information subtests of the Wechsler-Bellevue Scale, the four tasks were presented to each subject individually in a single session. Since it was desired to determine the cumulative effect of the variables related to the length of the tasks, it was necessary to control for position and for the possible effects of one task upon another. The orders of presentation of the four tasks were completely rotated, yielding 24 different orders. Thus every task appeared in each of the four positions, and followed directly each of the other tasks, an equal number of times. The alternate forms of each task also were rotated to eliminate differences in achievement resulting from possible inequality of the two forms.

In order to determine the subjects'

maximal levels of performance, failure could not be eliminated completely. A condition of minimal frustration could be obtained most effectively by making the number of successive failures as few as possible. The two forms of each task were administered, each in ascending order of difficulty. In the Digit Span task each form was terminated after three successive failures (a series length one digit beyond the just-determined memory span) instead of six as in the main experiment. In the Auditory Perception task each form was terminated after the subject could no longer identify any syllables in two successive steps. As in the main experiment, all of the Benton Visual Retention Test cards were presented and the Visual Perception task was continued until three successive failures had occurred. While the subjects were urged to guess when omissions were made, they were not informed of failure. Those subjects who appeared disturbed by failure were assured they had made a good try.

4. Results. No significant differences were found between the mean scores of the forms administered first and those administrated second. The t values obtained ranged from 0.56 to 1.06. (With an n of 23, t must be greater than 1.70 to be significant at the 0.10 level of confidence.) Pearson correlation coefficients between the first and second forms of each task ranged from 0.68 to 0.87. Although the size of the sample might tend to increase the possibility of negative results, no trend suggestive of a decrease in performance is evident between the forms administrated first and those administrated second. Three tasks showed slight increases in the scores of the forms administrated second; the Digit Span task showed a slight drop. The results are consistent with the assumption that any differences between forms obtained in the main experimental groups would be due to the introduction of the experimental variable.

In order to determine the effects of the position of the task on the initial performance of each of the four tasks and on the difference scores (form administered first minus form administered second), the raw scores in each task distribution were converted to standard scores. Each subject's scores then were arranged in a row according to the position in which the task was administered regardless of the nature of the task. An analysis of variance technique was employed to discover if any differences due to the position of the task existed. For the initial form scores of each of the four positions the analysis of variance technique yielded a *between-positions* variance ratio of 1.06, which is not significant. In addition, examination of the mean initial scores for the four positions revealed no consistent trends. For the difference scores the *between-positions* variance ratio also was not significant (1.18); the mean difference scores revealed no consistent trends. Thus the results are consistent with the assumption that any obtained positional effects in the main experimental groups are due to the cumulative effects of frustration.

C. MAIN EXPERIMENT

1. Population. There were two groups of subjects, a schizophrenic group and a normal group. The psychotic group was composed of 48 white male patients diagnosed as schizophrenic at the Veterans Administration Hospital, Northport, New York. They had been hospitalized for periods ranging from one to nine years, with a mean length of hospitalization of 4.75 years.

The normal group was composed of 48

white male patients hospitalized for minor surgery at the Veterans Administration Hospital, Bronx, New York. The following criteria of normality were employed: (*a*) hospitalization for relatively minor operations; (*b*) recommendation by the ward personnel as a suitable candidate on the basis of the absence of obvious behavior deviations; (*c*) no indication of previous mental disturbance nor of any present need for psychiatric treatment in the patient's medical history; (*d*) a screening interview by the experimenter prior to testing, in which the patient was questioned about previous illnesses, basis for discharge from service, and reasons for disability pensions, if any. Any patient whose mental status was doubtful was not accepted as a subject.

The groups were equated approximately for age and for intellectual and educational level. All subjects came from New York or immediately neighboring states. Ages ranged from 22 to 40 years, with a mean age of 30.77 for the schizophrenics and 30.00 for the normals. The estimated mean IQ's for the psychotic group and the normal group were 108 and 110 respectively (prorated from the Vocabulary and Information subtests), with a range of 86 to 128. The subjects had from 6 to 16 years of schooling, a mean of 11.31 years for the schizophrenics and a mean of 10.90 years for the normals.

A minimal intellectual level of Dull Normal ability (an IQ of 80) had been established to eliminate severely deteriorated psychotic patients. While the experimental battery was not administered to those individuals who failed to attain this level, the minimal need requirements (successful performance in the initial items of each task to obtain a score equal to three completely correct responses) also tended to disqualify the more confused and less compliant schizophrenic patients.

2. *Procedure.* All subjects were told that the psychologist was in the process of devising some new tests of different abilities, and that it would be appreciated if they would help him out by taking the tests. Most of the schizophrenic patients had received psychological examinations previously and usually submitted without further explanation. The surgical patients were unfamiliar with psychological tests for the most part and required more elaborate explanations. The exact wording of the introductory interviews varied with the patients' educational backgrounds and personality characteristics. Essentially they were told that the psychologist was working on a research project designed to compare the abilities of mentally ill and normal people through the use of some new tests. Normals were needed to form the basis for the comparison. All questions concerning the basis for selection and the tasks involved were answered as honestly as possible without revealing the true nature of the experimental variable.

The administration of the experimental battery to the main experimental groups was the same as the administration to the preliminary group except for the addition of the experimental variable as described earlier in this chapter. (See A. Materials.) The orders of presentation of the tasks were completely rotated twice within each group and the forms of each task also were alternated.

The failures encountered in the test situation tended to disturb the subjects in varying degrees. Primarily as a basis for reassurance, a brief interview (developed after five schizophrenic patients had been tested) was conducted with each subject at the end of the experimental session. The subjects were asked the following questions:

a. What do you think of these tests?
b. How do you think you made out?
c. How did you feel when you failed or got something wrong?
d. Which was the easiest and which was the hardest test?

If a subject's responses indicated that he was disturbed, he was assured that it was natural to feel upset, but that he had performed as well as or better than most people in his educational level. It was pointed out that psychological tests were analogous to the high jump or pole vault in sports. The athlete continues to jump until he misses several times. Unless he does miss, one could not know how high he might have gone. Because these were new tests, it was necessary to make certain that the subjects had reached their highest level. It also was suggested that the items that had been failed in the most difficult test were practically impossible.

II. RESULTS

A. Quantitative Differences in Frustration Tolerance

1. Drop Scores. For each task *drop scores* (i.e., the *postfrustration* score subtracted from the *prefrustration* score) were employed as measures of the adequacy of the subjects' performance following frustration. A large drop score indicates a low frustration tolerance.

The results clearly indicate that differences between schizophrenics and normals exist in their ability to withstand frustration. In all tasks the schizophrenic group showed significant mean drop scores (beyond the .001 level of confidence). On the other hand, the normal group manifested no reduction in the adequacy of performance following frustration in the perceptual tasks. The mean drop scores of the normals in the two perceptual tasks are well within chance limits. In the two memory tasks the normal group displayed drops in performance after frustration that were significant at the .035 level of confidence for the Digit Span task and at the .08 level of confidence for the Benton.

Comparisons of the mean drop scores of the two groups reveal that in all tasks the schizophrenic group consistently manifested a greater quantitative reduction in performance following frustration than the normal group did. The consistently poorer performance of the schizophrenic group as compared with the normal group following frustration tends to support the hypothesis that schizophrenics have a lower frustration tolerance than normals. The drop score means, standard deviations, and critical ratios of the two groups for each task, and the comparisons of the mean drop scores, are presented in Table 1.

2. Prefrustration Performance. Despite the approximate equality of the two groups in regard to Vocabulary and Information scores and educational levels, differences in the prefrustration mean

TABLE 1

Drop Score Means, Standard Deviations, and Critical Ratios of the Schizophrenic (S) and Normal (N) Groups for Each Task, and Comparison of Means of the Two Groups

	Digit Span		Benton		Aud. Per.		Vis. Per.	
	S	N	S	N	S	N	S	N
Mean	2.98	.67	2.98	1.12	4.92	1.08	10.38	.35
SD	2.73	2.20	4.11	4.42	6.04	4.72	13.39	9.84
CR	7.64	2.09	4.88	1.75	5.66	1.59	5.38	.25
CR	4.51		2.13		3.46		4.18	
P level	<.001		.033		<.001		<.001	

scores occurred in the Visual Memory and Visual Perception tasks. It is in these tasks that appreciable correlations between the prefrustration scores and the drop scores exist. For the Benton the Pearson correlation coefficients for the normal and schizophrenic groups are 0.52 and 0.29 respectively; for the Visual Perception task 0.37 and 0.43 respectively. These correlations probably are due largely to the unreliability of measurement. Extreme prefrustration scores are likely to include a considerable element of chance error which may not occur in the subsequent (postfrustration) trials with consequent regression toward the mean (the higher the prefrustration score, the bigger the drop score). In the two auditory tasks the correlations between prefrustration performance and drop score are considerably smaller and insignificant (0.20 or less).[1]

The obtained differences between the mean drop scores of the two groups in the visual tasks, therefore, are affected by the differences in prefrustration performance scores between the groups. In the Benton the fact that the schizophrenic group obtained a lower prefrustration mean score implies that, had the psychotic group attained the same prefrustration level as the normal group, the schizophrenics' mean drop score probably would have been larger than the obtained drop score. Statistically controlling for the difference in prefrustration performance by means of an analysis of covariance technique yields adjusted mean drop scores of 3.66 for the schizophrenic group and 0.38 for the normal group. The significance of the difference between groups then is increased to beyond the .001 level of confidence. (F equals 13.92; n_1/n_2 equals 1/93.)

In the Visual Perception task the difference between the initial mean scores of the two groups is reversed; the schizophrenic group obtained a somewhat higher prefrustration mean score. An analysis of covariance technique yields adjusted mean drop scores of 9.38 for the schizophrenic group and 1.50 for the normal group. While the difference between drop score means is reduced, the significance of the difference is still beyond the .001 level of confidence. (F equals 11.97; n_1/n_2 equals 1/93.)

B. QUALITATIVE DIFFERENCES IN REACTION TO FRUSTRATION

Various writers have emphasized different types of reaction to frustration. The Yale group (3) dealt exclusively with aggressive reactions, recognizing that they largely ignored consequences of frustration other than aggression. Symonds (13) treated somewhat more extensively these other consequences, such as constructive methods and substitute responses. In the latter category he listed abandoning the goal, repeating behavior, fantasy, regression, and miscellaneous reactions. Maier's experiments (7) were concerned with stereotyped, fixated behavior, but he also recognized the existence of other responses such as aggression, regression, and resignation (the lack of overt behavior).

The present experimental design was such that not all the different methods of reacting to frustration were elicited. For example, there was no clear-cut evidence of behavior that could be classified as regressive. In addition the relatively mild stress employed to induce frustration generally did not disrupt behavior to an ex-

---

[1] It is of interest to note that in the auditory tasks the physical differences between successive steps were equal (an increase or decrease of one digit in the memory task and of three decibels in the perceptual task). In the visual tasks the intervals between successive steps probably were more variable, which may have contributed to the greater unreliability of measurement.

TABLE 2

P LEVELS[1] AND *DF*S FOR THE QUALITATIVE DIFFERENCES BETWEEN GROUPS

	Before Frustration		During Frustration		After Frustration	
	P level	*df*	*P* level	*df*	*P* level	*df*
Aud. Per. Omissions	.70	3	.98	3	.02	3
Vis. Per. Omissions	.70	1	.20	3	.01	2
Aud. Per. Guesses	.90	3	.01	3	.80	3
Perseverative Errors	.20	1	.001	2	.05	2
Vis. Per. Response Time	.30	2	.70	3	.05	1

[1] Only the larger of the two *P*-level values between which the chi-square value falls is given in this table.

treme degree. The kinds of reactions that did occur tended to be less spectacular than those obtained by other experimenters under conditions of greater stress.

The quantitative reduction in the postfrustration scores was a result of an increase in errors. For the most part errors consisted of either totally incorrect responses (guesses) or refusals to respond (omissions). Some understanding of the basis for the difference in frustration tolerance between the schizophrenics and the normals may be obtained by analyzing the types of errors made.

A chi-square technique was employed to compare the two groups in the perceptual tasks with respect to such qualitative factors as the occurrence of omissions, guesses, and length of time for responses. The two perceptual tasks were divided into three parts for the comparisons: (*a*) the *before frustration* section in which subjects in both groups experienced mostly success; (*b*) the *during frustration* section in which all subjects experienced failure; (*c*) the *after frustration* section which was equivalent in item difficulty to the before frustration section. The *P* levels and degrees of freedom of the chi-square tables for the qualitative variables are presented in Table 2. For the after frustration section the *P* levels reflecting the difference between the groups were calculated on the basis of the change in the number of omissions, number of guesses, number of perseverative errors, and time scores for each individual (number in the before frustration section subtracted from the number in the after frustration section). The change rather than the absolute count is a more accurate measure of the effect of the intervening frustration.

1. Abandoning the Task. A tendency to abandon a task can be inferred from the change in the number of omissions made following frustration. An *omission* means that a subject refused to attempt to identify any part of a stimulus in the perceptual tasks. It should be recognized that a refusal to respond when the goal is unattainable is not necessarily an inadequate mode of behavior. The individual who correctly evaluates a situation as impossible may be justified in abandoning the task. Nothing would be gained by expending energy in a fruitless search for a response for which sufficient guiding cues are unavailable. However, to continue to abandon the task when success is possible is clearly inadequate behavior.

In the Auditory Perception task most subjects in both groups (28 normals and 27 schizophrenics) made one or less omissions during the before frustration section (steps 1 through 6 of the prefrustration form). When success became unlikely during the six most difficult steps, omissions increased, but the number of individuals in each group refusing to respond to one or more items remained approximately the same. The majority of the subjects (29 normals and 27 schizophrenics) made eight or more omissions during frustration. In the after frustration section (steps 6 through 1 of the postfrustration form), however, 29 schizophrenics (60%) increased the number of omissions, whereas only 18 normals (38%) did so. (See Table 2 for the *P* values of the obtained differences between groups.)

In the Visual Perception task no significant differences between groups were manifested before frustration (those items prior to the three consecutive failures). Almost all subjects (43 normals and 41 schizophrenics) made no complete refusals to respond. During frustration (the six most difficult items) omissions increased with all but nine normals and fourteen schizophrenics making one or more omissions. After frustration (those items in the postfrustration form equivalent in difficulty to the items in the before frustration section) fifteen schizophrenic subjects (31%) showed an increase of two or more omissions, whereas only four normals (8%) showed an increase of as many as two omissions. (See Table 2 for *P* values.)

The results indicate that it was only after frustration, when the goal again became attainable (as judged by each subject's level of performance prior to frustration), that differences between the groups became evident. More schizophrenic than normal individuals increased the number of omissions after frustration. Many schizophrenic subjects were unable to offer responses to items of a level of difficulty in which they had been successful prior to the frustrating experience.

A tendency to abandon a task can also be inferred from the time spent in attempting to discover a correct response. Each subject's average time for response to the blurred words was calculated for each section of the Visual Perception task. These final response-time scores showed a great deal of variation within as well as among individuals. On the less blurred words responses were made rapidly by most subjects, whereas on the more difficult items some subjects kept on trying to identify a word for as long as two minutes. In the before frustration section more schizophrenics (23%) than normals (10%) displayed extremely long reaction times (twenty or more seconds), but the difference between groups is not significant. No differences were evident during frustration. After frustration 22 normals (46%) decreased their average time spent per card, whereas 32 schizophrenics (67%) decreased their average response times. (See Table 2 for *P* values.)

These results are in accord with the findings regarding omissions. Thus, one of the factors contributing to the lower frustration tolerance in the schizophrenic group is their tendency to withdraw from the situation after having encountered frustration.

2. *Perseverative Behavior*. In the before frustration section of the Auditory Perception task the number of subjects in each group making one or more completely incorrect attempts to respond (guesses) was approximately the same (29 normals and 26 schizophrenics). In the after frustration section there also was no significant difference between groups in the number of subjects increasing the number of guesses. The major portion of both groups (30 normals and 28 schizophrenics) made as many or fewer guesses in the after frustration section than in the before frustration section. During frustration, however, when virtually no

useful perceptual cues were present, the schizophrenic group made significantly more guesses than the normal group. In this most difficult section of the task, 11 normals made from four to eight guesses, whereas 28 schizophrenics made from four to thirteen guesses. (See Table 2 for *P* values.)

Examination of the nature of the guesses revealed that some guesses appeared to be perseverative errors consisting in part of words or syllables that were direct repetitions of responses offered earlier in the test. In addition the records show conceptual perseverative guesses—the perseveration of an idea rather than a specific syllable or word (e.g., "baseball, football, soccer," or "piano, Mendelssohn, flute" offered in immediate succession). In the former example, "football" was classified as a perseverative error because of the incorrect repetition of the syllable "ball." This response also could be classified as a conceptual perseverative error along with the immediately following term "soccer." In the second example only the last two words were classified as perseverative guesses. The initial incorrect response "piano," which apparently provided the cue for the later responses, was classified as a nonperseverative guess. If one syllable of a two-syllable response correctly identified part of the stimulus word, the response was not classed as a guess despite the fact that the incorrect syllable might have been a perseveration.

Although somewhat more individuals in the schizophrenic group (25%) than in the normal group (15%) made perseverative errors in the before frustration section, the difference between groups is not significant. During frustration the difference was increased markedly with 33 schizophrenics (69%) and 18 normals (38%) making such errors. After frustration 15 schizophrenics (31%) and only six normals (12%) manifested an increase in the number of perseverative guesses. (See Table 2 for *P* values.)

Twelve schizophrenic patients made from four to eleven perseverative attempts, whereas only one normal made more than four perseverative guesses. In several of these extreme cases the behavior during and following frustration resembled in some degree the fixated, stereotyped behavior that Maier (7) and others (8, 13) have described.

While perseverative attempts did not appear to be directed at the goal which the experimenter had established, they cannot be considered samples of unmotivated behavior. It is important to recognize that the subjects had been told to guess when difficulty was encountered. The excessive number of perseverative responses made by some schizophrenic subjects suggests that they probably were complying mainly with the instructions to guess; i.e., merely to offer a response had been substituted for the original goal of identifying the spoken word. Many schizophrenics continued to respond in this perseverative manner even when the spoken words were loud enough to be identified (as determined by their performance prior to frustration). From the results obtained, however, it appears that for the vast majority of the normals the instructions to guess were not substituted for the instructions to identify the actual stimulus.

In regard to guesses in general it is of interest to compare the results in the Auditory Perception task with the findings of the Postman and Bruner study (9). Using an experimental and a control group of normals in a visual-perception task, Postman and Bruner subjected the experimental group to a large number of successive failures accompanied by "bitingly disruptive criticisms." The frustrated normals made more guesses including obviously inadequate attempts (meaningless phrases) than the control group.

The stress situation used by Postman and Bruner appears to have been much more severe than the one employed in the present experiment. During the mild failure situation the normal group made relatively few guesses. The schizophrenic group, however, reacted much more severely to the similar stress, making many more guesses, including the obviously inadequate, perseverative attempts. These findings suggest that the responses of schizophrenics to relatively mild stress may resemble those of normals to severe frustration.

3. Mode of Incorrect Response. Although omissions and guesses were errors which contributed to the reduction of scores following failure, to make an attempt to respond when the goal is again attainable should be a more desirable form of behavior than to refuse to try. The relative adequacy of the two forms of behavior can be inferred from the relationship between the mode of incorrect response after frustration and the size of the drop score. In the after frustration section of the Auditory Perception task 25 normals and 29 schizophrenics made more omissions than guesses (*omitting mode*) obtaining mean drop scores of 3.00 and 4.66 respectively. The 23 normals and 19 schizophrenics who made as many or more guesses than omissions (*guessing mode*) obtained mean drop scores of —1.00 and 5.00 respectively. In the normal group a biserial correlation coefficient of 0.60 (sigma $r_b = 0.13$) exists between the omitting mode and the size of the drop score. This suggests that for the normals to make no attempt rather than guess after frustration has been encountered is a less adequate mode of behavior. No relationship is evident in the schizophrenic group where the biserial correlation is -0.04 (sigma $r_b = 0.18$). Apparently it made little difference whether the schizophrenic patients chose to omit or to guess, since many of their guesses were inadequate (perseverative) attempts.

4. Bizarre Behavior. On the whole a relatively small number of the schizophrenic subjects and none of the normals manifested any form of bizarre behavior that could be related to the experimental frustration.

a. Nonsense Word Responses. Language disturbances are frequent concomitants of schizophrenic disorders. One phase of such disturbances is the use of words that have meaning only for the individual. In the Visual Perception task many normals as well as schizophrenics offered spelled responses that did not resemble meaningful words. Some subjects spontaneously commented that the response was not a real word; others admitted that the response was not a real word after being asked to define it. A meaningless response was classed as a nonsense word only when the subject defined it in some manner or stated that he believed it to be a real word. The following are examples of nonsense words given by schizophrenic patients:

Fronise—confronting or facing.
Enicely—infringing upon property.
Osflaze—a process applied to poetry.
Altrslt—a Harvard University graduation.
Figture—like an ashtray.

Responses that were only slight distortions of real words and defined accordingly were not classed as nonsense words. The response "Oculary," defined as "something to do with the eyes," was apparently a distortion of the word "ocular." Similarly, the response "Oswasso," defined as "a town in Michigan," was presumably a distortion of Owosso which is a town in that state.

In the schizophrenic group 13 subjects gave a total of 85 nonsense word responses, and defined most of them. Two normals gave a total of three words that were meaningless, and only one word was actually defined. Both normals had obtained very low Vocabulary scores.

It had been hypothesized that schizophrenics would resort readily to the use of nonsense words during and after frustration. The 85 nonsense words, however, were distributed fairly evenly among the before, during, and after frustration sections of the task (26, 31, and 28 nonsense words respectively) suggesting that responding with nonsense words was not a direct result of the experimental frustration. The use of nonsense words for these subjects appeared to be an habitual, uncritical means of dealing with the environment.

As in the case of perseverative guesses, offering nonsense words may reflect attempts on the part of the subject to comply with the instructions to guess. Seven (54%) of the 13 schizophrenic subjects who gave such meaningless responses made no omissions in this task. Of the 35 psychotic subjects who were more critical of their responses, a considerably smaller percentage (17%) made no omissions.

The patients making such meaningless responses also appear to have been less adversely affected by the frustrating experience than the psychotic individuals who were more critical of their responses. The tendency to give nonsense words is related negatively to the size of the drop score in the schizophrenic group. The biserial correlation coefficient between the presence of nonsense words and the size of the drop score is -0.46, with a standard error of 0.16.

b. Hallucinations. Hallucinations were observed in three schizophrenic patients during the experimental sessions. In one instance the hallucinations appeared to be induced by the failures in the final task.[2] For another patient the hallucinations resulted in a complete withdrawal from the reality situation. When the spoken words became too low to identify, he began to carry on what appeared to be a radio conversation and was oblivious to the actual task thereafter. A third subject merely reported seeing "a purple band of light while the record was playing," but there is little evidence to suggest that this hallucination was a consequence of frustration.

c. Fantasy. The use of fantasy to deny frustration may be inferred from the statements of four psychotic patients at the conclusion of the testing situation. Three of these subjects denied that they had failed or that they had encountered difficulty with any of the items in the tasks. One patient reported that he had made "a hundred per cent" in all the tasks.

It appears then that some forms of bizarre behavior (such as hallucinations and fantasy) may occur as a consequence of relatively mild frustration in a small number of schizophrenic patients. Other forms of bizarre behavior (such as the use of meaningless words) are more likely to be habitual means of dealing with the environment.

5. Report of Feelings. The descriptions of feelings, obtained in response to the question on how the subject felt when he had failed, were classified into two major types. One fairly clear-cut category included responses which indicated that the subject recognized and *accepted* unpleasant feelings. The other category, considerably broader, consisted of varied *defensive* reactions. Included in the defensive reactions were all verbalizations in which the subjects directly denied any painful experiences, or in which they avoided discussing their feelings. The avoidance of discussion of feelings was accomplished in essentially three ways: (*a*) denial of ego involvement in the tasks, thus implying that they were not disturbed in any way; (*b*) mere description of the failure situation; and (*c*) denial that they had failed (exclusively a schizophrenic mechanism).

---

[2] The following are excerpts from his remarks after encountering failure in the Digit Span task: "I give up. I can't do the test. I'm jamming my brain. . . . Lousy IQ. I hate you, you bastard. So you got a higher IQ. . . . He's jamming me with the radio. Turn it off. Let me listen to the psychologist. . . ."

It was generally quite difficult for most subjects to express their feelings in words. The responses of all the normal subjects, however, were classifiable. Several schizophrenic subjects appeared to be bewildered by the question; they apparently did not understand what was expected or could not describe their feelings. Other than the repetition of the question, no pressure was put upon the individual to respond. Hence, "I don't know" replies were not classified.

It seems that for the vast majority of the subjects the terms "failed" and "wrong" in conjunction with feelings connoted unpleasantness, since most of the replies to the question, including denials, referred to specific unpleasant emotions. Examples of the types of responses given by the normals (N) and the schizophrenics (S) are listed below.

Acceptance of Unpleasant Feelings:

I mean —. I knew —. How can I feel — a little stupid. (N)

I felt that I should have studied more. I felt more willing. I wanted to concentrate. I wanted to open my ears even though you can't. (N)

Depressed naturally, they were simple enough. (N)

On the numbers I didn't mind too much. On the others I felt a little ashamed. Gee whiz! They're so easy you know — for a guy who graduated from high school. (N)

I felt hurt. I said to myself I must be stupid. (N)

You feel miserable. I tried harder. (S)

I didn't — just badly. I felt like weeping when I failed continuously, made one (mistake) after the other. (S)

I have nervous heart trouble. (S)

Badly, I wanted to find myself in better condition. (S)

I can't stand it; that's the truth. I'm trying. (S)

Defensive Reactions:

It didn't annoy me. (N)

I didn't feel badly about it at all. (S)

I haven't got anything to gain or lose so there's no reason for feeling bad. (N)

Well it didn't matter to me whether I got it right or wrong because I wasn't taking the test for a reason. (S)

I got it wrong. I didn't know it that's all. (N)

Just went to the next one. (S)

I never fail. (S)

I couldn't be wrong. You've got to be right or commit suicide. (S)

In reporting these statements the emotional qualities in the subject's speech and manner cannot be conveyed. In addition the introspective reports, especially the defensive reactions, do not reflect the individual's actual behavior in many cases. For example, a normal, who denied any concern about his failures, returned several days later to ask for his score. A schizophrenic subject, who reported that he didn't "give a damn" if he passed or failed, was so severely disturbed during failure items in one task that hallucinatory behavior occurred.

In the schizophrenic group[3] 58 per cent of the subjects whose responses could

---

[3] A standard set of questions was developed after the first five subjects in the schizophrenic group had been tested. In order to increase the number of subjects, the interview data of nine additional schizophrenic patients were pooled with the data of the experimental group. These additional subjects had been tested to provide a source of substitutes for equating the two experimental groups.

TABLE 3

FREQUENCY OF TYPES OF RESPONSES TO INTERVIEW QUESTIONS

Interview Question	Type of Response	Normal $N=48$	Schiz. $N=57$
How did you feel when you failed or got something wrong?	Acceptance	32	20
	Defensive	16	28
	"Don't Know"		3
	No Reply		1
	Not Questioned		5
How do you think you made out?	Average or Above	18	27
	Below Average	29	17
	"Don't Know"	1	5
	No Reply		3
	Not Questioned		5
Which was the easiest test?	Digit Span	2	8
	Benton	26	15
	Auditory Perception	17	19
	Visual Perception	3	6
	No Decision		6
	Not Questioned		3
Which was the hardest test?	Digit Span	24	17
	Benton	2	11
	Auditory Perception	2	4
	Visual Perception	20	16
	No Decision		6
	Not Questioned		3

be classified tended to react defensively. In the normal group only 33 per cent reacted defensively. The numbers of subjects in each category are shown in Table 3. A chi-square test of the difference between groups in regard to the number of individuals in the *acceptance* and *defensive* categories is significant between the .01 and .02 levels of confidence.

6. *Estimate of Level of Performance.* A subject's opinion of his performance was obtained from his response to the question: "How do you think you made out?" The self-evaluations in the schizophrenic group ranged from complete denial of the reality situation ("a hundred per cent in all of them") to admission of poor performance ("I think I made out bad"). The normals were less grandiose in their estimates, which ranged from a high of "about average, maybe a little above," to "pretty crummy."

In order to permit comparison of the responses of the two groups, the estimates were divided into two categories: *average or above* included statements which suggested that the individuals believed they had made an "average" or a "fair" or at least a "not too bad" score, *below average* included all statements which suggested that the subjects were critical of their performance.

More schizophrenic subjects (61 per cent of those whose responses could be classified) than normals (38 per cent) estimated their performance at the average or above level. A chi-square test of the difference between groups in regard to the number of individuals in the two categories is significant between the .02 and .05 levels of confidence. The number of subjects in each category is shown in Table 3.

7. *Difficulty of the Tasks.* Only a few subjects explained the basis of their ratings of task difficulty. Their comments suggest that the number of failures encountered was probably a major factor

in influencing the ratings. In several instances, as the following comments indicate, task difficulty was judged by comparing the number of failures in the different tasks:

> They seemed equally hard, but I think I missed more on the numbers.
>
> I only missed out on about three drawings so they would be the easiest.

Other comments suggested that the subject had anticipated better performance (i.e., fewer failures) in that particular task because of his specific needs:

> I know I have very good eyesight, and I don't know why I couldn't read them (the blurred words).
>
> The numbers for me . . . I felt it would be rather easy to remember them.

The ratings of the four tasks are shown in Table 3. The Digit Span and Visual Perception tasks appear to be the most difficult; the Benton and Auditory Perceptions tasks were rated most frequently as the easiest. The groups are in close agreement except for the Benton, which was rated as most difficult by 11 schizophrenics but by only 2 normals. The fact that the schizophrenic group performed more poorly in this task than the normal group tends to add credence to the judgments.

8. Aggressive Reactions. During the testing situation several schizophrenic patients made hostile comments. One psychotic individual, the only subject who refused to continue with the tasks after encountering failure, was frankly critical of the examiner. Via his hallucinations one schizophrenic directed his aggression at the "voice." In between failures a third patient warned the psychologist to "watch out for the pen. You might get it stuck; it's sharp." Among the normals there were no obvious expressions of aggression directed at the examiner. During the experimental session their comments consisted mainly of self-criticisms and rationalizations for failure. Even at the conclusion of the experimental battery, when asked for their opinion of the tests, direct hostile responses were rare. Three normals described the tests as "silly" or "stupid." One psychotic subject thought the tests were "very dead, dull"; another refused to continue with the tasks directing hostile remarks at the experimenter. A third psychotic patient initially expressed his approval, but apparently projected his criticism adding: "For someone who didn't understand, he would think they were foolish."

In general the replies to the question, "What did you think of the tests," were so varied in content that it was not possible to categorize the responses and to compare the groups with respect to them. For the most part the opinions seemed to be of a positive but somewhat evasive nature. Many subjects in both groups thought that the tests were "all right" or "interesting." Among the normals such statements were frequently qualified by stating that they really didn't know the purpose of the tests. Others tended merely to describe the tasks as "memory" or "intelligence tests," or commented on the difficulty without offering any opinions.

After encountering the large number of failures that these persons had, it appears highly improbable that the majority of the subjects in both groups truly could have felt that the tests were very good or interesting, or that (as one psychotic put it) "I enjoyed them very much."

Apparently the experimental situation was not conducive to the outward expression of hostile or aggressive feelings. The fact that the experimenter probably represented an authority figure to the patients may have, contributed to the inhibition of such feelings. The effect of the presence of the experimenter on the inhibition of the expression of anger is demonstrated in an experiment by Dembo, which Koffka (6) has described.

In a situation where aggression is inappropriate Katz (5) points out that "masking, or camouflaging behavior (increased politeness, affected indifference, etc.) serve to conceal from the experimenter the vulnerability of the thwarted person: his emotional tension, ego injury (feelings of inferiority, shame, wounded pride, etc.), and punishable impulses (such as aggression)."

In the opinions of the tests expressed by both the schizophrenics and the normals, masking behavior seemed to predominate. However, most of the normals did not attempt to conceal emotional tension or ego injury when questioned about their feelings during failure and their estimates of their performance. On the other hand, despite their grossly inadequate performance after frustration, most of the schizophrenic patients attempted to conceal emotional disturbance.

The subjects who accepted unpleasant feelings appeared to be blaming themselves for their failures. Under conditions of failure due to lack of ability, Rosenzweig (11) states that inwardly directed aggression is warranted and can be considered an adequate reaction to frustration. The present experimental conditions were such that the subjects could be expected to see themselves as instrumental in their failures. Thus, the more frequent reports of acceptance of self-blame among the normals indicate that in this respect, too, they reacted more adequately to frustration than the schizophrenics.

C. Cumulative Effect of Frustration

In order to determine the cumulative effect of frustration on the prefrustration scores of each of the four tasks and also on the drop scores, the raw scores in each task distribution were converted to *T* scores with a mean of fifty and a standard deviation of ten. Each subject's *T* scores then were arranged in a row according to the position in which the task was administered regardless of the nature of the task. An analysis of variance technique was employed to discover if any differences due to the position of the task existed in either the prefrustration scores or the drop scores of either group. The converted mean scores are presented in Table 4. Since the distributions of the two groups were treated independently, the differences between them are not reflected in the converted means.

For both groups the *between positions* variance ratios of the prefrustration scores and the drop scores were statistically insignificant. It was only in the prefrustration scores of the normal group that the variance ratio was above unity (1.10). Inspection of the converted mean

TABLE 4

Converted Mean Scores (*T* Scores) in the Four Positions

	Group	1st Position	2nd Position	3rd Position	4th Position
Prefrustration Performance Scores	Normal	51.73	49.77	50.15	48.12
	Schizophrenic	50.27	50.06	50.31	49.65
Drop Scores	Normal	51.40	48.85	49.71	50.15
	Schizophrenic	50.15	48.25	51.02	50.77

scores, however, reveals the suggestion of a trend towards a decrease in the normal group's prefrustration performances as the number of tasks increased. Since a cumulative effect was anticipated, a direct comparison between the prefrustration scores in the first and final (fourth) positions is justified. In the normal group the *t* test for the difference between means in the first and fourth positions approaches significance at the .08 level of confidence. In the schizophrenic group the difference between the extreme positions of the prefrustration scores is well within chance limits; the four prefrustration scores remained fairly constant.

Although the results concerning the cumulative effect of successive frustrating experiences are not decisive, they do cast doubt upon the hypothesis that schizophrenics show progressive effects of frustration that are greater than those manifested by normals. As the number of separate tasks in which frustration was encountered increased, a depressing effect on the prefrustration performances of the normals (but not on the prefrustration performances of the schizophrenics) was noted. This may reflect a reluctance on the part of some normals to engage wholeheartedly in tasks where failure seems inevitable, suggesting a decrease in their motivational level.

Jenkins (4) believes that schizophrenics have undergone a greater number of frustrations in real life than have normals, basing his reasoning on studies which have shown that schizophrenics as a group have had little success in school, many frustrations in family life, etc. It appears likely that as a result schizophrenics have developed defenses against frustration such as withdrawal or abandoning the task, fantasy to deny failure experiences, and perseverative behavior, which serve as tension-reducing devices. Hence, despite the grossly inadequate nature of such mechanisms in coping with the reality situation, they may protect the individual to some extent from the emotionally disruptive effects of frustrating experiences.

III. SUMMARY AND CONCLUSIONS

The purpose of this experiment was to measure and to compare the effects of frustration on the performance of schizophrenic and normal individuals. Two groups, 48 schizophrenic patients and 48 normals, were matched approximately for age, intelligence, and education. The individuals in both groups were subjected to frustration in four tasks (auditory memory, visual memory, auditory perception, and visual perception). Frustration was said to exist when the subject had demonstrated a need to attain success in a task, and success in certain subsequent items of the task was thwarted by virtue of the increasing difficulty of the objective situation.

Each task consisted of two equivalent forms; items within each form were graded in difficulty. When failure occurred, it was pointed out without critical comment. Measures of the subject's performance on the form administered prior to failure and on the equivalent, alternate form following failure were obtained. The difference between the two scores so obtained was employed as the measure of the adequacy of the individual's response to frustration.

Since attention and concentration difficulties are characteristics of schizophrenic behavior, it was necessary to determine if merely the length of the tasks would cause a decrement in the performance of the schizophrenics. A preliminary group of 24 schizophrenic subjects was subjected to the same tasks under conditions of minimal frustration. The number of successive failures was reduced and the subjects were not informed of errors. The results indicated that no change occurred in performance from the first to the second form of each task, and that no cumulative positional effects were evident. Thus, the results are consistent with the assumption that any changes in performance in the experimental groups may be attributed to the introduction of the experimental variable.

In the two memory tasks the mild stress situation was sufficient to reduce to some extent the scores of the normals, and, to a significantly greater degree, also those of the schizophrenics. In the perceptual tasks the normals were able to maintain their level of performance despite the intervening failures; the schizophrenics were not. Thus, the consistently poorer performance of the schizophrenic group following frustration as compared with the performance of the normal group supports the hypothesis that schizophrenics have a lower frustration tolerance than normals.

The schizophrenics also differed qualitatively from the normals in their reactions to frustration. After encountering frustration the schizophrenics manifested a greater tendency to abandon the tasks by refusing to respond and by spending less time with items. Although during the most difficult section of the Auditory Perception task the psychotic group made many more guesses than the normals, these guesses were to a large extent of a perseverative nature rather than genuinely fresh attempts at solution. After frustration, when the task was again solvable as judged by the subject's performance prior to failure, significantly more schizophrenics than normals made perseverative guesses.

Within the normal group those individuals who preferred to make an attempt to respond, even though incorrectly, rather than omit after having encountered frustration, tended to perform more adequately than those normals who made more omissions than guesses. For the schizophrenics, however, it made little difference whether they refused to offer responses or guessed, since many of their guesses were perseverative attempts. In respect to the greater number of guesses, including many inadequate attempts, the reaction of the schizophrenics to relatively mild stress resembles somewhat the behavior of severely frustrated normals.

Some forms of bizarre behavior (such as hallucinations and fantasy) occurred as a consequence of frustration in a small number of schizophrenic patients. Other forms of bizarre behavior (such as the use of nonsense words) appeared to be habitual means of dealing with the environment.

The normals were more ready to admit poor performance and the presence of unpleasant feelings (shame, depression, self-blame, etc.) accompanying failure than were the schizophrenics. Under conditions of failure due to lack of ability, self-criticism is probably a more adequate reaction to frustration than attempts to conceal emotional disturbance.

The experimental situation was not conducive to the overt expression of aggression. Hostile remarks were rare in both groups. Despite the large number of failures encountered, most subjects in both groups made polite comments of approval when asked for their opinions of the experimental tests.

As the number of tasks in which frustration was encountered increased, the normals manifested a trend in the direction of a progressive decrease in their prefrustration performances in the four tasks. The suggested cumulative effect of successive frustrations may reflect a lowered motivational level. Although the schizophrenics were less able to cope adequately with frustration within each task, their prefrustration scores in the four tasks remained fairly constant throughout. The inadequate reactions of the schizophrenics to frustration (such as withdrawal and perseveration) may serve as tension-reducing mechanisms so that continuing effects of frustration were not apparent.

BIBLIOGRAPHY

1. Benton, A. L. A visual retention test for clinical use. *Arch. Neurol. Psychiat.*, 1945, **54**, 212-216.
2. Brown, J. F. Reactions of psychiatric patients in a frustrating situation. *Bull. Menninger Clin.*, 1939, **3**, 44-64.
3. Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.
4. Jenkins, R. L. Nature of the schizophrenic process: A working hypothesis for therapy. *Arch. Neurol. Psychiat.*, 1950, **64**, 243-262.
5. Katz, I. Emotional expression in failure: A new hypothesis. *J. abnorm. soc. Psychol.*, 1950, **45**, 329-349.
6. Koffka, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.
7. Maier, N. R. F. *Frustration: The study of behavior without a goal.* New York: McGraw-Hill, 1949.
8. Patrick, J. R. Studies in rational behavior and emotional excitement. II. The effect of emotional excitement on rational behavior in human subjects. *J. comp. Psychol.*, 1934, **18**, 1-22.
9. Postman, L., & Bruner, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314-323.
10. Rickers-Ovsiankina, M. Studies in the personality structure of schizophrenic individuals: II. Reaction to interrupted tasks. *J. gen. Psychol.*, 1937, **16**, 179-196.
11. Rosenzweig, S. Frustration as an experimental problem. VI. A general outline of frustration. *Charact. & Pers.*, 1938, **7**, 151-160.
12. Shakow, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. Dis. Monogr. Ser.*, 1946, No. 70.
13. Symonds, P. M. *The dynamics of human adjustment.* New York: Appleton-Century, 1946.
14. Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words.* New York: Columbia Univer., 1944.
15. Wilensky, H. *The performance of schizophrenic and normal individuals following frustration.* Unpublished doctoral dissertation, New York University, 1951.
16. *Manual of instructions for auditory test No. 14.* National Defense Research Committee, Cambridge: Harvard Univer., March, 1945.

(*Accepted for publication, February 6, 1952*)

Vol. 66, No. 13 Whole No. 345, 1952

Psychological Monographs: General and Applied

Parents' Attitudes vs. Adolescent Hostility in the Determination of Adolescents' Sociopolitical Attitudes[1]

Kenneth Helfant[2]
The Rye Public Schools, Rye, New York

I. INTRODUCTION

The present study began with the belief that sociopolitical attitudes are the result of a complex interweaving of personality and social factors, but that probably in most cases social factors are the more important of the two.

To bring the scope of the study within practicable limits it was decided to deal with one particular age group—children of adolescent age. Young people of this age are striving for self-expression, but they are still strongly influenced by the ties of home and family.

The scope of the study was further limited by centering upon a comparison of two relationships: (*a*) the relationship between a particular social factor (parents' sociopolitical attitudes) and adolescents' sociopolitical attitudes; and (*b*) the relationship between a particular personality factor (adolescents' personal feelings of outwardly-directed hostility) and adolescents' sociopolitical attitudes. On the basis of existing studies in the field, it was decided that probably the most potent social force contributing to the sociopolitical attitudes of adolescents is the sociopolitical attitudes of their parents, and that probably the most potent personality factor influencing sociopolitical attitudes of adolescents is their personal feeling of outwardly-directed hostility, which finds expression, through the mechanism of displacement, in sociopolitical attitudes.

The particular sociopolitical attitudes selected for study were attitudes toward Russia, attitudes toward international relations in general, and attitudes toward war. Attitudes toward Russia were selected for study because relations with that country were at the time in the forefront of public discussion. For the

---

[1] I am very grateful to the many people who have assisted me in the planning and carrying out of this study.

Professors Goodwin Watson and S. Stansfeld Sargent were especially helpful in helping me to organize my thoughts and design the experiment. In the collection of the data the cooperation of the principals, teachers, students, and parents of Teaneck, Bergenfield, and Dumont High Schools in New Jersey was such as to give me added incentive to produce results worthy of their efforts in my behalf. In analyzing and writing up the results Professor Irving Lorge gave advice and stimulation without which I would have been unable to realize the full significance of the data.

Special thanks are due to Miss Anne McKillop for her repeated reading and rereading of the manuscript, providing me with a sounding board from which my ideas were reflected enriched by her constructive criticisms and cogent suggestions.

[2] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the faculty of Philosophy, Columbia University, 1951.

same reason, however, it was thought that attitudes toward Russia might represent a special case, and not adequately reflect more general underlying factors in attitude formation. For this reason, in addition to attitudes toward Russia, study was given to attitudes toward international relations in general, and toward war.

So limited, the study presents these hypotheses:

1. Adolescents' attitudes toward Russia, international relations, and war are significantly related to:

a. The attitudes of the adolescents' parents on the same topics.

b. The adolescents' personal feelings of outwardly-directed hostility.

2. Adolescents' attitudes toward Russia, international relations, and war are more closely related to the attitudes of the adolescents' parents on the same topics than to the adolescents' personal feelings of outwardly-directed hostility.

II. SUBJECTS, INSTRUMENTS, AND PROCEDURES

The over-all plan of this experiment was to give a battery of attitude scales and personality tests to a group of high-school seniors, and then to give the same battery of attitude scales (but not the personality tests) to their parents. The correlations between attitudes of parents and of children were then calculated, as were the correlations between the attitudes of children and the factor measured by the personality tests (personal feelings of outwardly-directed hostility).

A. Subjects

Teaneck High School, located in the homogeneous upper middle-class community of Teaneck, in northeastern New Jersey, was selected as the locale for this experiment. Nine regular senior classes in American history (212 students), judged by the high-school principal and the teachers of the classes to be representative of the entire senior class (350 students), were used as the original sample. Forty-six students were eliminated because their parents could not be reached or because of failure to complete the tests. The final sample consisted of the remaining 166 high-school seniors (84 boys and 82 girls) and their parents (165 mothers, and 165 fathers).

B. Instruments

The Attitude Scales

For the purposes of this study we accept Chein's definition of an attitude as "a disposition to evaluate certain objects, actions, and situations in certain ways" (1, p. 177).

1. *A Survey of Opinions and Beliefs about Russia.* Originally it was hoped that the G. H. Smith scale for the measurement of attitudes toward Russia (7) might be used for this study, but it was found, when the Smith scale was given to a pilot group, that there tended to be a piling up of scores at the lower (unfavorable toward Russia) end of the scale. A new scale was therefore constructed by the experimenter, according to the Likert technique of attitude-scale construction. The scale, which includes 7 items from the original Smith scale, is reproduced below.

As a check on internal consistency, in addition to the usual steps in scale construction according to the Likert technique, the final scale was subjected to an item analysis. The split-group method was used, and all items of the scale were found to differentiate the high from the low scoring group for each of the three

A SURVEY OF OPINIONS AND BELIEFS ABOUT RUSSIA*

DIRECTIONS: Indicate your feeling about each statement by putting a check (√) in the appropriate column. Be sure to put a check after every statement.

	Strongly Agree	Agree	Don't Know	Disagree	Strongly Disagree
1. In my opinion, the Russian leaders will honestly support a United Nations plan to keep the peace....................	5	4	3	2	1
2. I feel that at present Russia is the curse of mankind........................	1	2	3	4	5
3. It is my belief that Russia's ways of doing things are quite as good as ours....	5	4	3	2	1
4. I feel friendly toward Russia..........	5	4	3	2	1
5. The Russian form of government makes slaves out of people..................	1	2	3	4	5
6. I think that the U.S. should tell Russia that if they take over one more country we will declare war..................	1	2	3	4	5
7. It seems to me that it would be a good thing for this country if a few of the Russian ideas and practices were adopted by the U.S..................	5	4	3	2	1
8. Even in peacetime the Soviet leaders use force to hold the mass of Russian people in place—make them do things against their will..........................	1	2	3	4	5
9. I believe that Russia is a dictatorship worse than that of Hitler............	1	2	3	4	5
10. In my opinion Russia is better than most newspapers would make us believe.....	[illegible]	4	3	2	1
11. I have an unfavorable attitude toward Russia............................	1	2	3	4	5
12. It is my feeling that the Soviet Union is likely to start another European war, sooner or later......................	1	2	3	4	5
13. I believe that Russia sincerely wants to see the people in the small countries of Europe decide their own form of government............................	5	4	3	2	1
14. I like more things about Stalin than I dislike............................	5	4	3	2	1
15. I think that trying to get along with Russia is a waste of time.............	1	2	3	4	5
16. I believe that the Russian leaders are sincerely interested in the welfare of the common man......................	5	4	3	2	1

* The small numbers which appear in the boxes after each item were not in the scale when it was presented to the subjects of the experiment, but indicate the score given to a check in that box. A high score indicates a friendly attitude toward the variable being measured, while a low score indicates an unfriendly attitute toward the variable.

groups (students, mothers, and fathers). As a further check on validity, the students in the experimental group were asked to estimate their attitude toward Russia on a linear scale, and the correlation between self-ratings and score on the attitude scale was found to be .78.

The reliability of the scale was calculated by the split-half method. The correlation coefficients so obtained, when corrected by the Spearman-Brown prophecy formula, were .92 for the students, .86 for mothers, and .89 for fathers.

2. *Survey of Opinions and Beliefs*

about International Relations. An attitude of hostility in international relations is defined as being indicated by strong tendencies to criticize, reject, or coerce foreign countries or peoples, and feelings that we should take a more militant stand in international relations. A scale to measure such feelings was constructed by the experimenter according to the Likert technique of attitude-scale construction. It is reproduced below.

As a check on internal consistency, in addition to the usual steps in scale construction according to the Likert technique, the final scale was subjected to an item analysis. The split-group method

A SURVEY OF OPINIONS AND BELIEFS ABOUT INTERNATIONAL RELATIONS*

DIRECTIONS: Indicate your feeling about each statement by putting a check (√) in the appropriate column. Be sure to put a check after every statement.

	Strongly Agree	Agree	Don't Know	Disagree	Strongly Disagree
1. In my opinion, the United States should give up trying to be on friendly terms with other countries..................	1	2	3	4	5
2. I think that if the United States is friendly toward other countries they are not as likely to be aggressive toward us	5	4	3	2	1
3. In my opinion, only foolish dreamers believe that international friendliness can accomplish anything in the modern world...........................	1	2	3	4	5
4. I feel that in international relations it is just plain common sense to "love thy neighbor as thyself".................	5	4	3	2	1
5. I believe that the U.S. should send food and materials to any country that needs them.........................	5	4	3	2	1
6. In my opinion, we shouldn't risk our happiness and well-being by getting involved with other countries..........	1	2	3	4	5
7. I think that helping foreign countries is a waste of money..................	1	2	3	4	5
8. In my opinion, international good will is essential to the welfare of the United States........................	5	4	3	2	1
9. It is my belief that we should get even with any country that tries to take advantage of the United States..........	1	2	3	4	5
10. I feel that we can't have "peace on earth, good will to men," because other nations are not of good will................	1	2	3	4	5
11. I think that being friendly with other countries will do more good than harm...	5	4	3	2	1
12. It is my feeling that we should try to help all nations, whether we get anything special out of it or not..........	5	4	3	2	1
13. I think that other countries are always getting us into wars................	1	2	3	4	5
14. I think that being friendly with other nations is a real help in solving international problems.................	5	4	3	2	1
15. It is my belief that other nations are often plotting against us..............	1	2	3	4	5
16. In my opinion, all sensible people believe in trying to be friendly with other countries.........................	5	4	3	2	1

---

* See footnote to questionnaire on p. 3.

was used, and all items of the scale were found to differentiate the high from the low scoring group for each of the three groups. As a further check on validity, the students in the experimental group were asked to estimate their attitude toward international relations on a linear scale, and the correlation between self-ratings and score on the attitude scale was found to be .70.

The reliability of the scale was calculated by the split-half method. The correlation coefficients so obtained (corrected by the Spearman-Brown formula) were .84 for students, .71 for mothers, and .81 for fathers.

3. *Survey of Attitudes Toward War.* This scale consisted of the 16 most differentiating items from Likert's 18-item adaptation of the original Droba-Thurstone attitude-toward-war scale (2). The 16 most differentiating items were determined by giving the 18-item scale to a pilot group and conducting an item analysis of the results. The two items from Likert's adaptation which were eliminated because of their low differentiating power were Items 4 ("An organization of all nations is imperative to establish peace") and 6 ("War brings misery to millions who have no voice in its declaration").

The following are representative items from the scale: "The benefits of war outweigh its attendant evils," "In the supreme offer of his life on the battlefield, a man attains to the greatest heights of glory," "It is our duty to serve in a defensive war," "It is the moral duty of the individual to refuse to participate in any way in any war, no matter what the cause."

The reliability of the attitude-toward-war scale was determined by the split-half method (corrected by the Spearman-Brown formula) and found to be .77 for the students, .59 for the mothers, and .49 for the fathers.

4. *Intercorrelations Among Attitude-Scale Scores.* Table 1 presents the intercorrelations between the various scale scores for adolescents, mothers, and fathers respectively. In every case, a high score on the attitude scales indicated a friendly feeling toward the country, idea, or institution involved.

As can be seen from Table 1, the intercorrelations between the scales are all low. The intercorrelations are highest for the adolescents (absolute average

TABLE I

INTRAGROUP CORRELATIONS AMONG SCORES ON THE ATTITUDE SCALES

Group	Attitude	Friendliness in Int'l Rel's	Approval of War
Students	Approval of Russia	.30*	−.26*
	Friendliness in Int'l Rel's	...	−.29*
Mothers	Approval of Russia	.17*	−.10
	Friendliness in Int'l Rel's	...	−.05
Fathers	Approval of Russia	.25*	−.16*
	Friendliness in Int'l Rel's	...	−.05

* Significant at .05 level.

$r = .28$), next for the fathers (absolute average $r = .15$), and lowest for the mothers (absolute average $r = .11$). Adolescent attitudes of friendliness in international relations and approval of war are negatively related to a statistically significant degree ($r = -.29$), while the correlations between these two attitudes for mothers and fathers show that their attitudes on these topics are unrelated ($r = -.05$ in both cases). The differences between the correlations, attitude of friendliness in international relations vs. attitude of approval of war for mothers and fathers, as compared with that for adolescents, are significant at the .05 level. There are no gross discrepancies in the intercorrelations between the scales, inasmch as the signs of the correlation coefficientss are all what we would expect if the scales are measuring what they purport to measure.

Measures of Students' Personal Feelings of Outwardly-Directed Hostility

For the purposes of this study, personal feelings of outwardly-directed hostility are defined as strong tendencies to criticize, rebel against, coerce, or reject persons or things; feelings that one must fight to get along, that one would like to destroy or mutilate objects or persons in one's environment, that offense is a good defense. In an attempt to measure the amount of such feelings that students had, the following instruments were used: (*a*) a questionnaire of the semi-projective type (the Like-Dislike Scale), adapted from a part of the Interests and Activities Questionnaire used by the Progressive Education Association in their evaluation of the Eight Year Study; and (*b*) the Rosenzweig Picture-Frustration Study.

1. *The Study of Likes and Dislikes.* The 34 items from the original Interests and Activities Questionnaire of the Progressive Education Association (6) which were designed to measure aggressive thoughts and actions were interspersed with 66 other items which were simply fillers. Instead of asking students to mark the items on a three-point scale (Like—?—Dislike), as in the original questionnaire, a five-point scale (Like very much—Like—?—Dislike—Dislike very much) was used.

Following are some representative items which when marked "Like very much" (L+) or "Like" (L) were considered to indicate feelings of outwardly-directed hostility: "Throwing spit balls," "Watching a good fight," "Picking somebody's argument to pieces," "Looking at pictures of death and destruction."

The odd-even reliability coefficient for the relevant items of the Like-Dislike Scale was found to be .88 (corrected by Spearman-Brown formula).

2. *Rosenzweig Picture-Frustration Study.* This test was first described by Rosenzweig in 1945 (4). It consists of 24 cartoons depicting people in mildly frustrating situations and requires the testee to fill in the "balloon" above the frustrated person.

The following are some typical situations depicted in the cartoons to which the subjects are asked to respond: (*a*) The driver of an automobile is apologizing to a pedestrian for having splashed the latter's clothing (cartoon #1), (*b*) A customer is complaining to a clerk that she has brought back a new watch three times now because it refuses to go (cartoon #5), (*c*) A woman is apologizing to her partner in a card game for having made a stupid play (cartoon #15), (*d*) A man returning a torn newspaper explains apologetically that the baby caused the damage (cartoon #24). The instructio[illegible] for the test are as follows: "Each of [illegible] following pictures contains two or more people. One person is always shown talking to another. You are asked to write in the empty space the very *first* reply that comes into your mind. Avoid being humorous. Work as fast as you can."

The test booklets were scored according to the directions given by Rosenzweig *et al.* (5), but only the extrapunitive score was used for the purposes of this study. Rosenzweig describes extrapunitive responses as follows (3): "Extrapunitive responses are those in which the individual aggressively attributes the frustration to external persons or things. The associated emotions are anger and resentment."

The correlation between scores on the Like-Dislike Scale and the extrapunitive scores was calculated and found to be .39, indicating a statistically significant, though not high, degree of community between the two instruments.

Measure of Student Intelligence

Because previous studies have found a positive relationship between intelligence and internationalist point of view, it was thought that a measure of intelligence should be included in order to be certain that this relationship was not confounding the relationships under study. A 20-word vocabulary test adapted by R. L. Thorndike from the I.E.R. Intelligence Scale: C.A.V.D. (8) was therefore administered to all student subjects.

Measures of Certain Background Factors

The Parent Interview. In order to obtain further qualitative and quantitative information about the subjects of this experiment, a parent-interview form was constructed by the experimenter for use in interviews with parents of the adolescents. The parent-interview form was designed to give measures of: (*a*) amount of conversation about international relations that goes on in the home in the presence of the children: (*b*) amount of parental contact with other countries; (*c*) amount of parental and adolescent church attendance; (*d*) parental attitude toward the attitude scales; and (*e*) parental economic position.[3] Items (*a*), (*b*), (*c*), and (*e*) above were assigned numerical indices on the basis of the answers given to the relevant questions in each case.[4]

A quantitative measure of the reliability of the parents' answers to the questions on the parent-interview form was not obtained. The subjective impression of the experimenter and of the other interviewers was that parents had no difficulty or resistance in accurately answering the questions.

C. PROCEDURE

The students were given all the scales and tests listed above (with the exception, of course, of the parent-interview form) during their regular classes in American history. Two full class periods of 50 minutes each were required to complete the testing of the students. All student testing was done by the experimenter. The students were urged to fill out all scales and tests as frankly as possible, and assured that the results would not affect their school grades or records in any way. In order to further insure frank answers, students were not required to put their names on any of the scales or tests, but they were identified by means of a coding system.

---

[3] A copy of the parent-interview form and an explanation of the scoring system appear in Appendix A.

[4] The replies to item (*d*) were categorized and appear in Appendix B.

The parents of the students were interviewed individually in their homes. During the course of the interview they were given the attitude scales, which they filled out themselves in the presence of the interviewer. After the attitude scales had been filled out and interviewers had checked them to see that all items were answered, the parents were asked the questions on the parent-interview form, the interviewer writing down the answers. The interviewing was done by the experimenter and three graduate students from the Department of Sociology, Graduate Faculties, Columbia University.

The experimenter met with the interviewers before the actual interviewing began and explained the purposes and methods of the study in detail. A standard method for conducting the interviews was agreed upon, and while the interviewing was going on, periodic meetings were held to discuss any problems that came up. The experimenter went along to at least one of the interviews conducted by each interviewer to be certain that the agreed-upon standardized procedure was being followed.

III. ANALYSIS OF THE DATA

Because the hypotheses which this study was designed to test (see Ch. I) involve measurement of degree as well as kinds of relationships, it was decided that a correlational method of analysis was most appropriate. Table 2 presents the intercorrelations and the means and standard deviations of the scores on each of the major tests and scales administered to the 166 high-school seniors and their parents (165 mothers and 165 fathers).

The differences among the means of

TABLE 2

INTERCORRELATIONS, MEAN SCORES, AND STANDARD DEVIATIONS OF THE MAJOR VARIABLES

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Student approval of Russia	...	.28*	.18*	.30	.04	.13	−.26*	−.20*	−.06	−.05	−.02
2. Mothers' approval of Russia	...	...	.37*	.23*	.17*	.18*	−.13	−.10	.04	.02	.01
3. Fathers' approval of Russia		...	...	.22*	.16*	.25*	−.13*	−.13	−.16*	.05	.02
4. Student attitude of friendliness in Int'l Rel's	...	...	...	...	.11	.27*	−.29*	−.16*	−.04	−.18*	−.25*
5 Mothers' attitude of friendliness in Int'l Rel's	...	...	...	...	...	.32*	−.03	−.05	.01	.04	−.10
6. Fathers' attitude of friendliness in Int'l Rel's	...	...	...	...	...	...	−.13	−.11	−.05	.03	−.06
7. Student approval of war	...	...	...	...	...	...	...	.16*	.05	.16*	.12
8 Mothers' approval of war	...	...	...	...	...	...	...	...	.16*	−.02	−.01
9. Fathers' approval of war	...	...	...	...	...	...	...	...	...	−.06	.01
10. Student hostility (Like-Dislike)	...	...	...	...	...	...	...	...	...	...	.39*
11. Student hostility (Rosenzweig)	...	...	...	...	...	...	...	...	...	...	...
Mean scores	36.5	37.5	35.6	58.4	56.5	56.8	47.2	48.3	49.4	88.2	53.5
Standard deviations	10.0	7.8	9.2	8.1	6.6	8.4	8.2	5.1	5.2	17.6	15.4

* Correlations which are significant at the .05 level of confidence (which has been adopted as the level of statistical significance for this study) are marked with an asterisk.

the scores of parents and adolescents are not statistically significant for any of the attitude scores. The differences among the standard deviations of the attitude scores of students and their mothers are statistically significant[5] for all three of the attitudes measured. The differences between the standard deviations of the students' and fathers' scores on the attitude-toward-war scale, and the differences between the standard deviations of the mothers' and fathers' scores on the attitude-toward-international-relations scale are also statistically significant. This is to say that, in this sampling, mothers as a group showed a more homogeneous attitude toward Russia than their children, a more homogeneous attitude toward international relations than either their children or their husbands, and that both mothers and fathers showed more homogeneous attitudes toward war than their children.

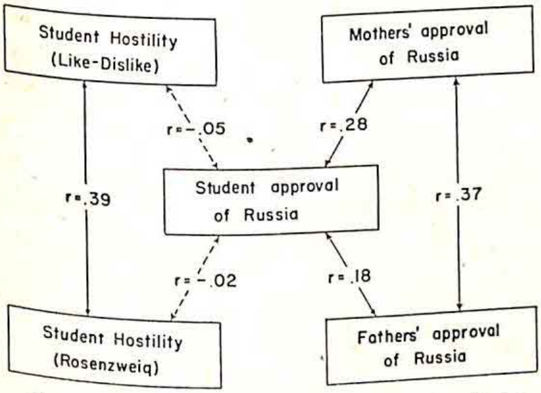

Student Hostility (Like-Dislike)
Mothers' approval of Russia
r=-.05
r=.28
r=.39
Student approval of Russia
r=.37
r=-.02
r=.18
Student Hostility (Rosenzweiq)
Fathers' approval of Russia

FIG. 1.—Attitude of approval of Russia. (Intercorrelations among student attitudes of approval of Russia, parents' attitudes of approval of Russia, and student personal feelings of outwardly-directed hostility.)

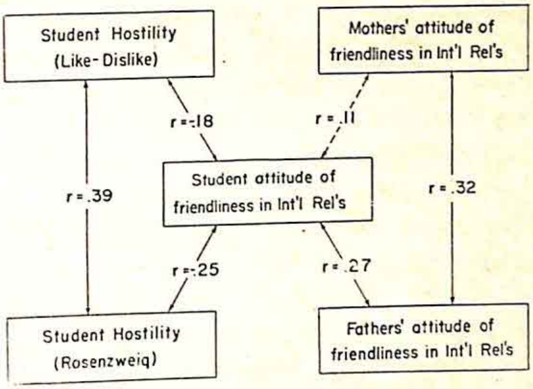

Student Hostility (Like-Dislike)
Mothers' attitude of friendliness in Int'l Rel's
r=-.18
r=.11
r=.39
Student attitude of friendliness in Int'l Rel's
r=.32
r=-.25
r=.27
Student Hostility (Rosenzweiq)
Fathers' attitude of friendliness in Int'l Rel's

FIG. 2.—Attitude of friendliness in international relations. (Intercorrelations among student attitudes of friendliness in international relations, parents' attitudes of friendliness in international relations, and student personal feelings of outwardly-directed hostility.)

Figures 1 through 3 illustrate the relationships among the major variables. In every case, a high score on the measures of hostility indicated hostile feelings, while a high score on the measures of attitudes indicated a friendly attitude toward the country, idea, or institution involved. A *solid line* joining two variables indicates that the correlation between the two variables is *significant at the .05 level.* A *dotted line* joining two variables indicates that the correlation between the two variables is *not statisti-*

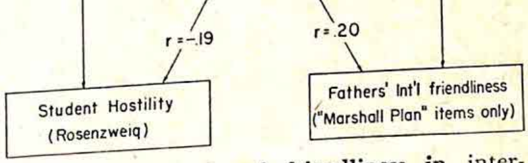

Student Hostility (Like-Dislike)
Mothers' Int'l friendliness ("Marshall Plan" items only)
r=-.04
r=.17
r=.39
Student Int'l friendliness ("Marshall Plan" items only)
r=.32
r=-.19
r=.20
Student Hostility (Rosenzweiq)
Fathers' Int'l friendliness ("Marshall Plan" items only)

FIG. 2*a*.—Attitude of friendliness in international relations: "Marshall Plan" items only. (Intercorrelations among student attitudes of friendliness on "Marshall Plan" items of international-relations scale, parents' attitudes of friendliness on these items, and student personal feelings of outwardly-directed hostility.)

---

[5] This was determined by the F test, where $F = \frac{s_2^2}{s_1^2}$.

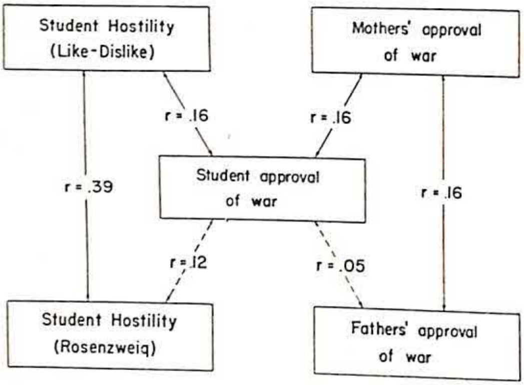


FIG. 2*b*.—Attitude of friendliness in international relations: "Ideological" items only. (Intercorrelations among student attitudes of friendliness on "Ideological" items of international-relations scale, parents' attitudes of friendliness on these items, and student personal feelings of outwardly-directed hostility.)

cally significant. In these figures and in the discussion which follows, when we speak of student hostility or of student or parent attitude, we are referring to scores on the instruments used for measuring hostility and attitudes.

The average of the correlations, student approval of Russia vs. mothers' approval of Russia ($r = .28$), and student approval of Russia vs. fathers' approval of Russia ($r = .18$), is .23. The average of the correlations, student approval of Russia vs. student hostility score on the Like-Dislike test ($r = -.05$), and student approval of Russia vs. student hostility scores on the Rosenzweig test ($r = -.02$), is −.03. The difference between these two average correlation coefficients is statistically significant.[6]

Because of the possibility that different items from the international relations scale might have been interpreted differently, in spite of the precautions taken against this when the scale was constructed, it was thought that it might be fruitful to examine separately the relationships between student and parent attitudes, and student hostility and student attitudes, for separate groups of items from the scale. Accordingly, two groups of items which were thought to represent different ways of looking at international relations were scored separately for each of the three experimental groups (students, mothers, and fathers), and these subscores intercorrelated among experimental groups and with the measures of student hostility. The results of these analyses are given in Figures 2*a* and 2*b*.

Figure 2*a* shows the intercorrelations that existed for scores on Items 5 ("I believe that the U.S. should send food and materials to any country that needs them") plus 7 ("I think that helping foreign countries is a waste of money") plus 12 ("It is my feeling that we should try to help all nations, whether we get anything out of it or not"). It was

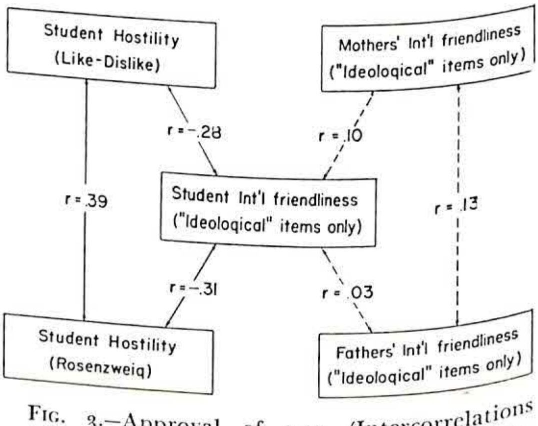

Student Hostility (Like-Dislike)
Mothers' Int'l friendliness ("Ideological" items only)
r = -.28
r = .10
Student Int'l friendliness ("Ideological" items only)
r = .39
r = .13
r = -.31
r = .03
Student Hostility (Rosenzweig)
Fathers' Int'l friendliness ("Ideological" items only)

FIG. 3.—Approval of war. (Intercorrelations among student attitudes of approval of war, parents' attitudes of approval of war, and student personal feelings of outwardly-directed hostility.)

---

[6] This was determined by converting the r's to Fisher's z's and using the formula

$$t = \frac{z_1 - z_2}{\sigma_{z_1 - z_2}}$$

where

$$\sigma_{z_1 - z_2} = \frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}.$$

thought that these three items, because they all refer to helping other countries, might be interpreted in terms of attitudes toward the Marshall Plan to a greater extent than others.

Figure 2*b* shows the intercorrelations that existed for scores on Items 1 ("In my opinion, the United States should give up trying to be on friendly terms with other countries") plus 11 ("I think that being friendly with other countries will do more good than harm") plus 14 ("I think that being friendly with other countries is a real help in solving international problems") plus 16 ("In my opinion all sensible people believe in trying to be friendly with other countries"). These four items were thought to represent an abstract ideological point of view to a greater extent than the other items of the scale.

The correlations in Figure 2*a* as compared to those in Figure 2*b* show that the correlations between student attitudes of friendliness in international relations and student hostility are higher for the "ideological" than for the "Marshall Plan" items of the international-relations scale, and that the correlations between parents' and students' attitudes are higher for the "Marshall Plan" than for the "ideological" items, but the differences are not statistically significant.

The means and standard deviations of the subscores intercorrelated in Figures 2*a* and 2*b* appear in Table 3.

As in the case of the differences between the means of the total scores of students and parents on the attitude-toward-international-relations scale, none of the differences between the means of the subscores of parents and students are statistically significant. The differences between the standard deviations of the subscores of students and their mothers are statistically significant in the case of scores on the "Marshall Plan" items. The differences between the standard deviations of the subscores of mothers and fathers are statistically significant in the case of scores on both the "Marshall Plan" and the "ideological" items. That is to say, mothers, as a group, showed more homogeneous attitudes on the "Marshall Plan" items of the attitude-toward-international-relations scale than their children. and had more homogeneous attitudes on both the "Marshall Plan" and "ideological" items of the international-relations scale than their husbands.

Because of the low reliability of the attitude-toward-war scale, it was thought that, as in the case of the attitude-toward-international-relations scale, it might be fruitful to examine the relationships under investigation separately for separate groups of items from the scale. Accordingly, two groups of items which were thought to represent different aspects of attitude toward war were

TABLE 3

MEANS AND STANDARD DEVIATIONS OF THE SUBSCORES FROM THE SCALE ON ATTITUDE TOWARD INTERNATIONAL RELATIONS

Subscore	Mean	Standard Deviation
Students' scores on "Marshall Plan" items	10.5	2.4
Mothers' scores on "Marshall Plan" items	10.5	2.0
Fathers' scores on "Marshall Plan" items	10.2	2.4
Students' scores on "ideological" items	16.6	2.1
Mothers' scores on "ideological" items	15.8	1.9
Fathers' scores on "ideological" items	15.9	2.2

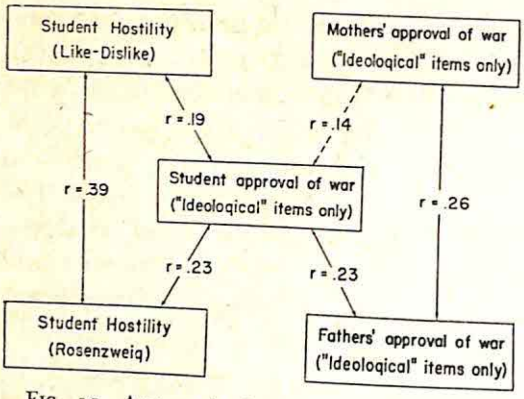

Student Hostility (Like-Dislike)
Mothers' approval of war ("Ideological" items only)
r = .19
r = .14
r = .39
Student approval of war ("Ideological" items only)
r = .26
r = .23
r = .23
Student Hostility (Rosenzweig)
Fathers' approval of war ("Ideological" items only)

FIG. 3a.—Approval of war: "Personal Participation" items only. (Intercorrelations among student attitudes of approval of war based only on "Personal-Participation" items of the attitude-toward-war scale, parents' attitudes of approval of war based on these items, and student personal feelings of outwardly-directed hostility.)

scored separately for each of the experimental groups (students, mothers, and fathers) and intercorrelated among experimental groups and with the measures of student hostility. The results of these analyses are given in Figures 3a and 3b.

Figure 3a shows the intercorrelations that existed for scores on the items "The benefits of war outweigh its attendant evils" plus "War in the modern world is as needless as it is suicidal" plus "There is no progress without war." It was thought that these three items represent an abstract ideological attitude toward war to a greater extent than the other items of the scale.

Figure 3b shows the intercorrelations that existed for scores on the items "If a man's country enters a war which he does not consider justified, he should nevertheless serve at the front or wherever he is needed" plus "It is our duty to serve in a defensive war" plus "He who refuses to fight is a true hero" plus "It is the moral duty of the individual to refuse to participate in any way in any war, no matter what the cause." It was thought that these four items represented an attitude toward personal participation or actual behavior in time of war to a greater extent than the other items of th[illegible]cale.

The correlations in Figure 3a as compared to those in Figure 3b show that the correlations between student attitudes of approval of war and student hostility are higher for the "ideological" than for the "personal-participation" items of the attitude-toward-war scale, and that the correlations between parents' and students' attitudes are of different signs for the "personal-participation" items as compared with the "ideological" items. The differences between the correlations student-father and father-mother for the "ideological" items of the war scale (Fig. 3a) and for the "personal-participation" items of the war scale (Fig. 3b) are statistically significant.

The means and standard deviations of the subscores intercorrelated in Figures 3a and 3b appear in Table 4.

As in the case of the differences between the means of total scores of students and parents

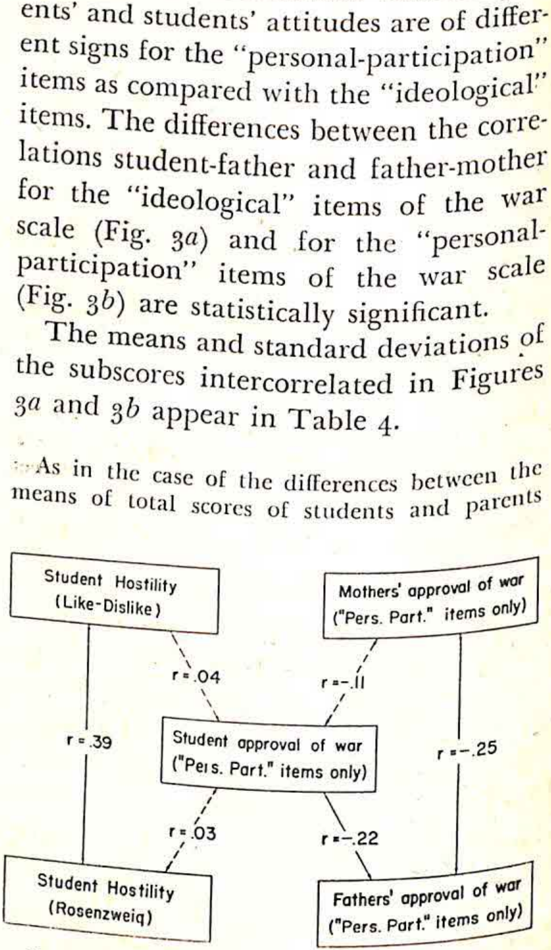

Student Hostility (Like-Dislike)
Mothers' approval of war ("Pers. Part." items only)
r = .04
r = -.11
r = .39
Student approval of war ("Pers. Part." items only)
r = -.25
r = .03
r = -.22
Student Hostility (Rosenzweig)
Fathers' approval of war ("Pers. Part." items only)

FIG. 3b.—Approval of War: "Ideological" items only. (Intercorrelations among student attitudes of approval of war based only on "Ideological" items of the attitude-toward-war scale, parents' attitudes of approval of war based on these items, and student personal feelings of outwardly-directed hostility.)

TABLE 4

MEANS AND STANDARD DEVIATIONS OF THE SUBSCORES FROM THE SCALE ON ATTITUDE TOWARD WAR

Subscore	Mean	Standard Deviation
Students' subscores on "ideological" items	5.7	2.4
Mothers' subscores on "ideological" items	6.2	2.0
Fathers' subscores on "ideological' items	6.1	2.0
Students' subscores on "personal-participation" items	16.6	2.7
Mothers' subscores on "personal-participation" items	16.2	0.6
Fathers' subscores on "personal-participation" items	16.8	1.8

on the attitude-toward-war scale, none of the differences between the means of the subscores of parents and students are statistically significant. The differences between the standard deviations of the subscores of mothers and their children and mothers and their husbands are statistically significant in the case of both the "ideological" and "personal-participation" items of the attitude-toward-war scale. The differences between the standard deviations of the subscores of fathers and their children are statistically significant in the case of the "personal-participation" items. This is to say, that mothers, as a group, had a more homogeneous attitude on both the "ideological" and "personal-participation" items of the attitude-toward-war scale than either their children or their husbands, and that fathers, as a group, had a more homogeneous attitude on the "personal-participation" items of the attitude-toward-war scale than their children.

Effect of Home Conversation about International Relations

The correlations between indices of amount of home conversation about international relations and scores on the attitude tests were all small and lacked statistical significance. The scores were further analyzed by subgroups, because it was thought that this procedure might reveal a relationship which was being obscured or minimized by the Pearson correlation technique. Accordingly, separate correlations were calculated for the two extreme groups: parents who reported that they talked a great deal about international relations with their children ($N = 48$), and parents who reported that they did not talk at all about international relations with their children ($N = 39$). The results of this analysis showed that correlations between adolescents' and parents' attitudes were generally greater in homes where parents and children discussed international relations than in homes where they did not, but the increases in the correlations were not statistically significant.

In order to find whether there were any differences in the kind of attitudes held when there was a great deal of conversation about international relations in the home as compared to when there was none, the mean scores of adolescents, mothers, and fathers were calculated separately for these two extreme groups. The results of this comparison showed that the mean scores on the attitude tests were the same whether there was no conversation or a great deal of conversation about international relations in the home.

Effect of Parental Contact with Other Countries

Indices of amount of parental contact with other countries (obtained from answers to questions under Section 2 of the parent-interview form) were correlated with all other data, and found to be significantly correlated only with students' vocabulary-test scores ($r = .19$), and home-conversation scores ($r = .22$).

Because, as in the case of home conversation about international relations, it was thought that there might be a relationship between amount of parent contact with other countries and students' and parents' attitudes which was being masked by the Pearson correlational technique, it was thought advisable to compare the mean attitude scores of the two extreme groups in this respect. Parents who received a score of zero in "amount of contact with other countries" ($N = 12$) were compared with those who received the highest possible score ($N = 21$). No statistically significant differences were found. Thus, neither the comparison of extreme groups nor the correlational method of analyzing the data revealed any significant relationships between the attitudes studied and amount of parental contact with other countries.

Effect of Frequency of Church Attendance

Indices of the frequency of adolescents' church attendance (obtained from answers to Section 3 of the parent-interview form) were correlated with all other data, and found to be significantly correlated with student attitudes toward international relations ($r = -.22$).

Effect of Economic Position of Parents

Indices of the economic positions of the parents used in this study (obtained from answers to questions under Section 5 of the parent-interview form) were correlated with all other data, and found to be significantly correlated with student vocabulary scores ($r = .28$), and mothers' attitude-toward-war scores ($r = .21$).

Effect of Adolescent Intelligence

Adolescent vocabulary scores, which were taken as a measure of adolescent intelligence, were correlated with all the data collected, and found to be significantly correlated with adolescents' attitudes toward international relations ($r = .21$), attitudes toward war ($r = .20$), mothers' attitudes toward international relations ($r = .20$), and index of parental economic position ($r = .28$).

Because it was thought that the correlation between intelligence and adolescents' attitudes toward international relations and war might be confounding the correlations between adolescents' and parents' attitudes on these topics, the correlations between adolescents' and parents' attitudes on these topics were recalculated with vocabulary score partialed out. It was found that this procedure had a negligible effect on the adolescent-parent coefficients, and it was therefore concluded that though vocabulary score (intelligence) was correlated with attitudes toward war and toward international relations, it had no appreciable effect on the relationships under consideration.

IV. DISCUSSION

In general, it can be said that the original hypotheses of the study were supported in the case of some attitudes, but not in the case of others. The correlations are all small, even in the cases where the hypotheses are supported. From the results of this study as well as other studies, it is apparent that the relationships that exist between child and parent attitudes, as well as between attitudes and personality factors, are dependent upon many factors. In view of this fact, we would expect to find only small correlations between attitude and any single factor.

The specific findings of the study will be discussed under two headings: influence of parents' attitudes, and influence of personal feelings of outwardly-directed hostility.

Influence of Parents' Attitudes

From the results of this study parents appear to be more compartmentalized in their attitudes than are their children. When we examine the table of intercorrelations for the major variables (Table 2), we note that the variable, student attitude of friendliness in international relations, leads all others with respect to number of significant intercorrelations. It is significantly correlated with 8 of the principal variables of the study, while the other attitudes (student approval of war and of Russia) are significantly correlated with only 5 and 4 variables respectively. It seems likely that the reason for this is the fact that adolescents' attitudes toward international relations are derived to a considerable extent from parental discussions about specific aspects of international relations, for example, Russia and war. The parents' attitudes toward international relations, however, are not related to the same extent to specific aspects of such relations, such as war and our relations with Russia. We find that our adolescents' attitudes toward international relations are related to parental attitudes toward war and Russia, but that the adolescents' attitudes toward war and Russia are not related to parental attitudes toward international relations (Table 2). The fact that parents' attitudes, especially mothers', are not interrelated to the extent that adolescents' attitudes are, is brought out quite clearly in Table 1. The lower reliability of the attitude scales for the parents as compared to adolescents is a further indication that the parents are more compartmentalized in their thinking (i.e., their answers to the items of the scales show less item-to-item consistency than the answers of their adolescent offspring).

Another generalization that can be drawn from the data of this study is that adolescents tend to adopt an attitude which is somewhat more extreme or emphatic than the attitude of their parents, especially their mothers. The standard deviations of the attitude scores and subscores of the mothers are in every case smaller than those of the adolescents: significantly so for all scores and subscores except those on the "ideological" items of the international-relations scale. The standard deviations of the attitude scores and subscores of the fathers are in every case except scores and subscores on the attitude-toward-international-relations scale (where they are approximately equal) smaller than those of the adolescents: significantly so in the case of total scores on the attitude-toward-war scale and subscores on the "personal-participation" items of the attitude-toward-war scale. On examining the pattern of scores in individual cases, it appears that the reason for the greater variation in adolescent sociopolitical attitudes as compared to those of mothers and, to a lesser extent, fathers, might be that when adolescents' attitudes resemble those of their parents at all, they almost always have a more extreme score than that of their mothers (and often that of their fathers too), though in the same direction. That is, if an adolescent has a mother who has a mildly friendly attitude toward Russia, for example, the adolescent is likely to have a distinctly friendly attitude toward Russia, and contrariwise, if the mother has a mildly unfavorable attitude toward Russia, her adolescent offspring is likely to have a distinctly unfriendly attitude toward Russia. This "leverage" theory offers an explanation of the fact that there are no significant differences among the mean scores of the adolescents and parents, but frequent significant differences among the standard deviations of their scores.

The results of the analysis of the "ideological" as compared with the "personal-participation" items of the attitude-toward-war scale (Fig. 3*a* and 3*b*) bring to light the interesting fact that, for this population, parents and children tend to have similar ideological attitudes toward war, but tend to have dissimilar attitudes toward personal participation in war. In the case of the "ideological" items, adolescents' attitudes and those of their parents are positively correlated (significantly so in the case of adolescents and fathers, very near significance in the case of adolescents and mothers), but in the case of the "personal-participation" items adolescents' attitudes and those of their parents are *negatively* correlated (significantly in the case of fathers, nonsignificantly in the case of mothers). The average of the correlations students vs. mothers and students vs. fathers on the "ideological" items of the war scale is .18. The average of the correlations students vs. mothers and students vs. fathers on the "personal-participation" items of the war scale is −.16. The difference between these two average correlation coefficients is statistically significant.

Most of the correlations between parents' and students' attitudes and between attitudes of husbands and wives which are reported here are lower than those reported in other studies. This may be due to differences in samples studied, or it may be due to the fact that previous studies have depended upon the child subjects to obtain the attitudes of their parents. The sample studied here differs from others in that the children involved were exclusively adolescents within a narrow age group, and it might be said that the low degree of relationship between parent and child attitudes is a result of the adolescent revolt which caused subjects to reject the attitudes of their parents at this particular stage in their development. This does not, however, explain why the correlations between husbands and wives were lower than those usually reported, and this fact turns us toward the second possible explanation, that previous studies have depended upon the child subjects to obtain the attitudes of their parents. Since only subjects willing to obtain the attitudes of their parents were included, it is possible that in previous studies this may have had a biasing effect on the results. It seems likely that subjects who are willing to obtain the attitudes of their parents are probably more cooperative in general and have better relationships with their parents than those who are not. Such children could be expected to identify more closely with their parents, and one would expect that the relationships between their attitudes and those of their parents would be higher than those between children and parents in general. We would also expect parents to have a better-than-average relationship to each other in cases where they tend to have a better-than-average relationship with their children. Such parents would tend to identify with each other to a greater extent than in the general population, thereby bringing about a higher relationship between their attitudes than in the general population.

Influence of Personal Feelings of Outwardly-Directed Hostility

The problem that is posed by the results of this study is not whether there is a relationship between adolescents' personal feelings of outwardly-directed hostility and their sociopolitical attitudes, but why it is that relationship exists for some of the attitudes studied (international relations and war), but not for others (Russia). In other words, why is it that sometimes adolescents' attitudes show a statistically significant correlation with their personal feelings of outwardly-directed hostility, and sometimes do not? The results of this experiment suggest an hypothesis in answer to this question; it is necessary first, however, to review findings from the selected groups of items (or subscales) of the full scales.

Because of the low reliability of the attitude-toward-war scale, probably due to lack of homogeneity of the scale, the total scores on this scale have no clear-cut meaning, and it appears more reasonable to consider only the subscores on the "ideological" and "personal-participation" items of the scale. For the same reason, it appears more reasonable to consider only the "ideological" and "Marshall Plan" items of the international-relations scale. After thus limiting ourselves, we are left with five attitude scores: attitude toward Russia, scores on the "Marshall Plan," and "ideological" items of the international-relations scale, and scores on the "ideological" and "personal-participation" items of the attitude-toward-war scale. These scores may be arranged into two general categories: scores which reflect attitudes which are subject to strong social pressures (e.g., Russia, "Marshall Plan" items of the international-relations scale, and the "personal-participation" items of the attitude-toward-war scale), and scores which are more abstract and less subject to social pressures (e.g., the "ideological" items of the scales measuring attitude toward international relations and attitude toward war).

If we examine the figures showing the correlations between the scores of adolescents on the above measures and their scores on the tests of adolescents' personal feelings of outwardly-directed hostility (Fig. 1, 2*a*, 2*b*, 3*a*, and 3*b*), it appears that the highest correlations between attitude and the measures of hostility occur for the "ideological" items of the international relations and war scales, in regard to which there tends to be little social pressure. As a matter of fact, with only one exception there are no significant relationships between any of the more specific attitudes toward which there tends to be considerable social pressure, and the measures of hostility. On the other hand, if we examine the correlations between the above scores and parents' scores, it appears that the lowest correlations between adolescent attitudes and those of their parents are on the abstract "ideological" items, toward which there tends to be less social pressure; and the highest correlations between adolescent and parental scores tend to be on the more specific attitudes toward which there tends to be considerable social pressure. The hypothesis which these results suggest is that *when social pressures with regard to an attitude are strong*—as they are, for example, in the case of attitudes toward Russia—*the effect of personality factors is overruled; and, conversely, when social pres-*

sures with regard to an attitude are not strong, the attitude held is more likely to be consistent with the person's characteristic ways of reacting. This hypothesis offers an explanation for many attitudinal phenomena the explanation for which is at present not very satisfactory. It offers an explanation, for example, of why many Southern whites who do not seem to be displacing personal feelings of hostility nevertheless have strongly hostile attitudes toward Negroes: it may be because of the social pressure from their own group, which would ostracize them if they professed any but the (Southern) socially approved attitude.

The instruments used in this experiment were not designed to test the above hypothesis, and it seems likely that the exceptions to it that were found may be due to this fact. None of the so-called "Marshall Plan" items actually mention the Marshall Plan, and it seems likely that these items were interpreted in a general abstract way as well as in terms of the specific "Marshall Plan." This may be why there was a statistically significant correlation between adolescents' scores on these items and their scores on the Rosenzweig test. Similarly, the "ideological" items are not strictly "ideological" and free from social pressures, as we would like in order to test the hypothesis; and this may be why there is a statistically significant correlation between adolescents' and fathers' scores on the "ideological" items of the attitude-toward-war-scale. In addition to the fact that our instruments were not designed to test the above hypothesis, we have of necessity concerned ourselves with only a fragment of the social and personality forces acting on the individual.[7] To test the hypothesis adequately would take a different and more elaborate study than this one, but one which might add considerably to our knowledge of the way attitudes are formed and function in the life of the individual.

V. SUMMARY

The present study began with the belief that sociopolitical attitudes are the result of a complex interweaving of personality and social factors, but that probably, in most cases, social factors are the more important of the two. An experiment was conducted to test this belief by investigating the relationship among sociopolitical attitudes of 166 adolescents (high-school seniors), sociopolitical attitudes of their parents (165 mothers and 165 fathers), and some measures of adolescent hostility.

The sociopolitical attitudes investigated were (*a*) attitudes toward Russia, (*b*) attitudes toward international relations, and (*c*) attitudes toward war. The scales on attitude toward Russia and attitude toward international relations were constructed by the experimenter using the Likert technique; the scale on attitude toward war consisted of a 16-item Likert version of Droba's scale.

The measures of hostility consisted of the extrapunitive scores of the adolescents on the Rosenzweig Picture-Frustration Study, and their scores on a Like–Dislike scale of a semiprojective type, adapted by the experimenter from a scale used for the measurement of aggressive impulses by the Progressive Education Association in the evaluation of the Eight Year Study.

---

[7] As was stated in Section I, in planning this study, an attempt was made to select factors which seemed most important in determining the sociopolitical attitudes of adolescents. There is a question, however, as to whether each of the "most potent" factors affecting sociopolitical attitudes of adolescents which have been selected are equally "most potent." On the side of social forces, it seems likely that we have what is probably the "most potent" factor (parents' attitudes). In the personality sphere, however, we may have what is only "somewhat potent" rather than "most potent." A fuller measurement of personality factors might lead to higher correlations with this aspect.

Measures of adolescents' intelligence, amount of conversation about international relations that went on in their homes, amount of parental contact with other countries, frequency of church attendance, and parental economic position were obtained.

The attitude scales and personality tests were given to the adolescents during their regular classes in American history, and parents' attitudes were obtained by administering the attitude scales to the parents during the course of interviewing them in their homes. All variables were expressed in terms of numerical scores and intercorrelated.

The main conclusions of the study were that for the population studied:

Correlations between adolescents' attitudes and those of their parents are smaller than those indicated by previous studies. Previous studies have relied on adolescents themselves to obtain the attitudes of their parents, thus introducing a selective factor which may have resulted in spuriously high correlations.

There are small but statistically significant correlations between adolescents' and mothers' attitudes toward Russia and war ($r = .28$ and $r = .16$ respectively); between adolescents' and fathers' attitudes toward Russia ($r = .18$) and toward international relations ($r = .27$); between adolescents' attitudes of friendliness in international relations and both measures of adolescent hostility (average $r = -.26$); and between adolescents' attitudes of approval of war and hostility scores on the Like-Dislike Scale ($r = .16$).

Parents' attitudes (especially mothers') appear to be more compartmentalized than those of their adolescent offspring. The adolescents' attitudes studied were all significantly intercorrelated. The attitudes of fathers were significantly intercorrelated only in the case of attitudes toward Russia vs. attitudes toward international relations, and in the case of attitudes toward Russia vs. attitudes toward war. The attitudes of mothers were significantly intercorrelated only in the case of attitudes toward Russia vs. attitude toward international relations.

The influence of mothers' and (to a lesser extent) fathers' attitudes on those of their adolescent offspring tends to take the form of the adolescents adopting an attitude which is somewhat more extreme or emphatic than the attitude of their parents.

Adolescents' attitudes toward Russia are significantly more closely related to their parents' attitudes toward Russia than to their personal feelings of outwardly-directed hostility.

An analysis of subscores on groups of items from the scales on attitude toward international relations and attitude toward war suggest the hypothesis that *when social pressures with regard to an attitude are strong, the effect of personality factors is overruled; and, conversely, when social pressures with regard to an attitude are not strong, the attitude held is more likely to be consistent with the person's characteristic ways of reacting.*

Neither amount of conversation in the home about international relations nor amount of parental contact with other countries has a significant effect on the adolescents' attitudes studied nor the degree of relationship between adolescents' and parents' attitudes.

There were no statistically significant differences among mean attitude scores of adolescents and their parents, but the standard deviations of the scores show mothers as a group to have significantly more homogeneous attitudes than adolescents as a group.

Intelligence, as measured by the Thorndike 20-word vocabulary test is significantly correlated with attitudes toward international relations and war, but was found to have a negligible effect on the relationships under investigation.

For adolescents, frequent church attendance is reliably associated with less liberal attitudes toward international relations.

APPENDIX A

PARENT INTERVIEW FORM[1]

Name ....

Address .... Tel. No. ....

Score

1. Do you talk about international relations here at home?

(1) (———) never (2) (———) rarely (3) (———) sometimes
(4) (———) often

Amount of talk about international relations ......

2. Have you ever lived abroad? (———) yes (2)

Do you have friends who have lived abroad? (———) yes (1)
Do you speak any foreign languages? (———) yes (2)
Do you have friends who speak any foreign language? (———) yes (1)
Do you have friends who come from other countries? (———) yes (1)

Index of contact with other countries ......

3. Does your son or daughter attend church?

(1) (———) never (2) (———) rarely (3) (———) sometimes
(4) (———) often

Index of frequency of church attendance ......

4. What do you think of the questionnaires you have just taken?

5. Estimated cost of house: ———(1) under $10,000
———(2) $10,000 to 15,000
———(3) $15,000 to 20,000
———(4) $20,000 to 25,000
———(5) over $25,000

Estimated cost of furnishings ——— (rate on a 5 point scale, (1) very cheap, (3) average, (5) very expensive)

Do you own a car? (———) yes (1)

Do you have a telephone? (———) yes (1)

What is your profession (husband)?

(1)——— Unskilled worker, including laborer or domestic servant.
(2)——— Semi-skilled worker.
(3)——— Proprietor of small business (value less $5000).
(4)——— Skilled worker.
(5)——— Clerk or kindred worker.
(6)——— Semi-professional or smaller official of large business.
(7)——— Professional or proprietor of large business (value more than ($5000).

Index of economic position ......

Scoring of the Parent Interview Form

The scoring of the parent interview form consisted of simply adding up the weights assigned to an affirmative answer to each question under each of the four sections of the form for which quantitative indices were derived (Sections 1, 2, 3, and 5). The answers to Section 4 (parent reactions to the attitude scales) were written in by the interviewers and later categorized and tabulated by the experimenter. The tabulation of parental replies to this question appears in Appendix B.

---

[1] The categorization of professions used in this form is adapted from that of Warner, W. L., Meeker, M., & Eells, K. *Social class in America: A manual of procedure for the measurement of social status*. Chicago: Science Research Associates, 1947.

APPENDIX B

ADOLESCENT AND PARENT REACTIONS TO THE ATTITUDE SCALES

Parent reactions to the attitude scales were categorized and tabulated by the experimenter. The results of this tabulation appear in Table 5.

As can be seen from Table 5, most of the comments of the parents about the questionnaires were unfavorable.

On the basis of observation while interviewing, and after checking with the other interviewers, it is the opinion of the experimenter that the principal reason for the parents' unfavorable attitudes toward the scales was that many of the parents seemed to feel threatened by the scales, which represented a "test" to them. Although they were assured that the scales were only to get their opinions, and had no relation to their intelligence or the amount of information they possessed, they had never come in contact with scales of this sort before, and felt that they were some sort of a test or examination. The fact that the quesions on the scales are overlapping and contradictory made them feel that we were trying to catch them in inconsistencies. This "threat" aspect of the scales suggests that somewhat more accurate and open expressions of opinion might have been obtained by discussion of leading questions with the interviewers, who could then have rated the attitudes of the respondents. This of course would introduce new errors in the subjective nature of the ratings of the interviewers. If a careful semi-controlled interview scheme were worked out, however, it seems likely that this method might yield somewhat more accurate measurement of attitudes than the use of attitude scales with adult subjects of the type used in this study.

The adolescents were generally enthusiastic about the study and their part in it. In answer to Item 95 of the Likes and Dislikes questionnaire, "Answering questionnaires like this," 68 per cent of the adolescents responded by circling the "Like" or "Strongly Like" symbols, 17 per cent circled the question mark, and only 15 per cent circled the "Dislike" or "Strongly Dislike" symbols. Adolescents' attitudes toward the questionnaire were given a numerical index of from 1 to 5 on the basis of their replies to Question 95 of the Likes and Dislikes scale, and these indices were correlated with all other variables. No significant correlations were found.

TABLE 5

PARENT REACTIONS TO THE ATTITUDE SCALES

Comment	Number of of Times Commented Upon
Hard to answer because need to qualify	41
Overlapping questions	16
Ambiguous	14
Repetitious questions	11
Vague	10
Communistic	6
Drawn out	1
War propaganda	1
Shows type of citizen	1
Presumptuous	1
Loyalty test	1
Foolish	2
Satisfactory (O.K.)	16
Good	8
Thorough	5
No comment	29

REFERENCES

1. Chein, I. Behavior theory [illegible] the behavior of attitudes; some crit[illegible] comments. *Psychol. Rev.*, 1948, **55**, 1[illegible]88.
2. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.*, 1932, No. 140, 55 pp.
3. Rosenzweig, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944, Vol. I, ch. 11, pp. 379-388.
4. Rosenzweig, S. The picture-association method and its application in the study of reactions to frustration. *J. Pers.*, 1945, **14**, 3-23.
5. Rosenzweig, S., *et al.* Scoring samples of the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1946, **21**, 45-72.
6. Smith, E. R., Tyler, R. W., *et al. Adventure in American education.* Vol. III. *Appraising and recording student progress.* New York: Harper, 1942. 550 pp.
7. Smith, G. H. Attitudes toward Soviet Russia: I. The standardization of a scale and some distributions of scores. *J. soc. Psychol.*, 1946, **23**, 3-16.
8. Thorndike, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, **26**, 1942, 128-135.

Accepted for publication February 16, 1952.

Vol. 66, No. 14 Whole No. 346, 1952

Psychological Monographs: General and Applied

Personality of the Convulsive Patient in Military Service

T. W. Richards*
Louisiana State University

INTRODUCTION

During 1945, at the height of World War II, there came to the U. S. Naval Hospital at San Leandro, California, a large number of patients about whom there was question of diagnosis of "epilepsy." At the suggestion of the medical officer in command of the hospital, Dr. W. F. Kennedy, an intensive study of these patients was made.

Three broad lines of investigation were established. First, each patient was given the usual *medical* examination. The neurological examination was particularly thorough; prominent among the special techniques used was the water Pitressin test. A second examination was *electroencephalographic* (hereafter referred to as the EEG); under resting conditions and following hyperventilation the electrical activity of the brain was measured; records were judged "abnormal," "borderline," and "normal." (A report of these medical findings has been published previously (9). The third type of examination was *psychological,* consisting of several procedures. Two questionnaires filled out by the patient were checked in interview with him. The psychological instruments which were applied include:

a. The Cornell Selectee Index (CSI). This scale, designed for screening at induction and recruiting stations, consists of 64 items concerned for the most part with subjective physical complaints.

b. A history blank, covering briefly items regarding parents, schooling, jobs, arrests, health, accidents, and military experience.

c. The Minnesota Multiphasic Personality Inventory (MMPI) (8). (The standard procedure used for the scale was modified slightly by eliminating from the total of 550 items the 250 which do not contribute toward a score.) This inventory asks that the subject decide about each of a number of statements whether it is "True" of himself or "False," and it allows him to express uncertainty or indecision by classifying the item "Don't know" or "?". Main variables of the MMPI in its original standardization were determined empirically on the basis of differences between "normals" and groups classified according to neuropsychiatric diagnosis. Thus, a scale designated as *Hy* is composed of items which distinguish between controls and a group of hysteric patients. Of these main variables, there are the following: *Hs* (hypochondriacal patients); *D* (depressive patients); *Hy* (hysteric patients); *Pd* (psychopathic deviate patients); *Pa* (paranoid patients); *Sc* (schizophrenic patients); and *Ma* (manic patients). Also, there is a scale, *Mf* (scored for *interests* similar to those of the sex opposite to oneself).

It is arranged that the raw scores for each of the variables of the MMPI may (by transmutation into *t* scores, where the mean equals 50 and one sigma equals 10 points) be plotted on a chart. Any patient may be compared easily with the norms by means of this chart.

---

* In this research the author has utilized the services of several persons whose kindness he wishes to acknowledge. Basic test data were gathered in the hospital atmosphere by the author and three colleagues to whom he is indebted especially: Barbara H. Krieger, Suzanne H. Ray, and Dr. G. T. Lodge.

Because of their participation in the analysis of special sorts of data, the author wishes here to thank Paul Greenberg, Winifred Ingram, and Jean Paulsen. He wishes to acknowledge a debt in particular to Natalie Kettleman, whose dissertation (10) was concerned with the Rorschach data.

d. Parts of the Wechsler-Bellevue Intelligence Scale (W-B) (21). This included for most patients five tests of the Verbal Scale (Information, Comprehension, Arithmetic, Digits, and Similarities) and the Block Designs Test of the Performance Scale.

e. The Rorschach ink blot test, administered and scored according to the method of Klopfer (11).

f. The Thematic Apperception Test (TAT). This was used in a brief form involving ten pictures of the revised series of Morgan and Murray (13): No. 3BM, 4, 6BM, 7BM, 9BM, 12M, 13MF, 14, 17BM, and 18BM.

Supplementary to this routine battery of psychological tests, several other tests (the Healy Pictorial Completion II and the Shipley-Hartford Retreat Scale) were used selectively in individual cases.

The present report concerns the performance of the convulsive patient on the psychological tests, and the implications of the findings for the theory of personality in general and of the personality of the convulsive patient in particular.

It will be found that this report is organized about a conceptual scheme which emphasizes Capacity, Control, and Motivation. The concepts of these three general areas of behavior have been found useful in accounting for personality and its organization (17). The notion of *Capacity* is intended to include those aspects of the personality involved in productivity and potentiality and any discrepancy between the two. *Control* is suggested as a concept to include those methods of dealing with reality which the individual has developed through his unique experiences. This would involve much of what is known in other framework as ego psychology, for it has to do, so far as emotionality is concerned, with *methods of defense* against anxiety or threat. The third category, *Motivation*, refers to basic impulses—needs and threats which initiate the individual to adjust—to demonstrate capacity and control.

The group studied. The material presented here was obtained between January and August, 1945, from a group of 72 men in naval service. They ranged in age from 17 to 43 years, with a median of 23.4 years. Length of service (in the Navy or Marine Corps) varied from 6 months to 7 years, with a median of 2.2 years. Three patients had completed college; 24 (one-third) were high-school graduates. Only 11 men had less than full grade-school education. All but 2 men had been overseas, and more than one-half (54 per cent) had been in combat, almost exclusively in the Pacific area, as Marines (ground troops), or as sailors aboard ship. About one-half of the men (47 per cent) had been married at least once. In 5 cases (7 per cent) there was instance of either definite separation or divorce. The fact that these men had a record of active duty over a considerable period of time is important, since most studies of "epileptic" patients utilize hospitalized cases of long standing.

In response to the question, *Was your family broken up by divorce or death of parents before you were 16?*, 21 patients answered in the affirmative. This represents 29 per cent of the group. Of these 21 cases, 5 reported divorce or separation; 7 indicated death of the mother; and 9 death of the father. In 3 of these instances, the father had been killed.

Neurophysiological findings. Medical findings established the following points (9): (*a*) A history of seizures prior to enlistment was elicited in 47 per cent of the total group. The age of the first seizure ranged from infancy to 42 years.

(*b*) While 59 per cent of the group gave a history of head trauma of varying degrees of severity, the neurologist felt that "no individual in this series [illegible]d seizures which suggested very strong[illegible] a relationship to head trauma" (9, p. 2).

EEG study of the total group of 72 patients revealed tracings which to the encephalographer were "abnormal" in 32 cases (46 per cent), "borderline" in 19 (26 per cent), and "normal" in 21 (29 per cent). In summary, he may be quoted as follows:

> There was nothing consistent in the character of the abnormal EEG's. Most of them contained slow cycles of varying frequencies and about one-fourth had organized rapid cycles (frequencies between 18 to 30 cycles per second), usually with slow activity. In only 1 of the 73[1] cases was a typical spike and wave pattern obtained. Yet several of the patients had *petit mal* attacks, although all but three had *grand mal* seizures also (9, p. 3).

Relevant previous research. The report of Lisansky (12) in 1948 summarizes the research studies of the personality of convulsive patients. Test findings obtained by means of the Rorschach test by Rorschach (18), Guirdham (5), Stauder (19), Harrower-Erickson (7), and Arluck (1) fail in general to support the concept of an "epileptic personality." Lisansky's own research, which compared ten epileptics and ten diabetics matched roughly for age, educational achievement, age of onset, and duration of illness suggested that the epileptics had lower W-B performance IQ (99.5 compared with 105.7 for the diabetics). With the Rorschach their response time was longer and they demonstrated more of the "neurotic signs" of Klopfer, "more emotional strain and less acceptance of self than does the diabetic group . . . . On the whole, however, the groups show far more similarity than differences" (12, p. 37). Lisansky comments on the lack of studies of noninstitutionalized, nondeteriorated adult patients using projective personality test.

INTELLECTUAL CAPACITY OF THE CONVULSIVE PATIENT

It was pointed out that only 11 patients, or 15 per cent, had less than complete grade-school education. One-third had finished high school, and 3 men (4 per cent) were college graduates. This leaves a group of 53 per cent who had some high-school education, but did not graduate. A number of these men left school to enter the service. On the grounds of schooling alone, then, it is clear that we are dealing with a group which is of at least average intellectual competence.

Discussion of the results of performance on tests of mental ability will be confined for the most part to the Wechsler-Bellevue Scale (hereafter referred to as W-B), and more specifically, to six of the ten subtests of this Scale. These are the five verbal tests (Information, Comprehension, Digit Span, Arithmetic, and Similarities) and a single performance subtest, Block Designs. Furthermore, since the performance of only 59 of the total group of patients was sufficiently beyond suspicion to qualify it for research evaluation, this discussion will be limited to the results obtained with a portion (82 per cent) of the total group.

Intelligence quotients for the full scale (five verbal tests and block designs) ranged from 57 to 132, with a mean of 98.3, a standard deviation of 17.0. The mean is close to that for the general population, but variability for the convulsive group is greater, although the difference is not significant statistically.

---

[1] The electroencephalographer included in his total group one case additional to the group given psychological tests.

The pattern of performance. An analysis of the group results with subtests of the W-B indicates slight superiority of the performance (in this instance the Block Design test) over the verbal tests. However, the difference is not significant statistically. Examining this trend in somewhat greater detail, reference may be made to Figure 1. This represents for each subtest, in terms of weighted scores (whereby the normative mean is at 10 and sigma is set at 3 points), the central tendency and variation within one sigma for the standardization data furnished by Wechsler (21) on adults of comparable age, and for the group of convulsive patients.

It is seen that the convulsives differ from the controls most in their poorer performance with Arithmetic and with Digits; these are differences significant at the .01 level of confidence or better. The difference for Information (with the convulsives again being less apt) is at the .05 level. Differences for the remaining subtests—Comprehension, Similarities, and Block Designs—are negligible. Three patients achieved IQ scores below 70—namely, 57, 62, and 62. History in each case, together with supplemental testing, seemed to reinforce the impression of borderline mental deficiency. In two cases the EEG was considered "normal," in the third, "borderline."

The concept of "deterioration." In the literature on epileptics or convulsive phenomena there is considerable tendency to regard such behavior as reflection of "organic" defect or impairment. With this idea in mind, there was calculated

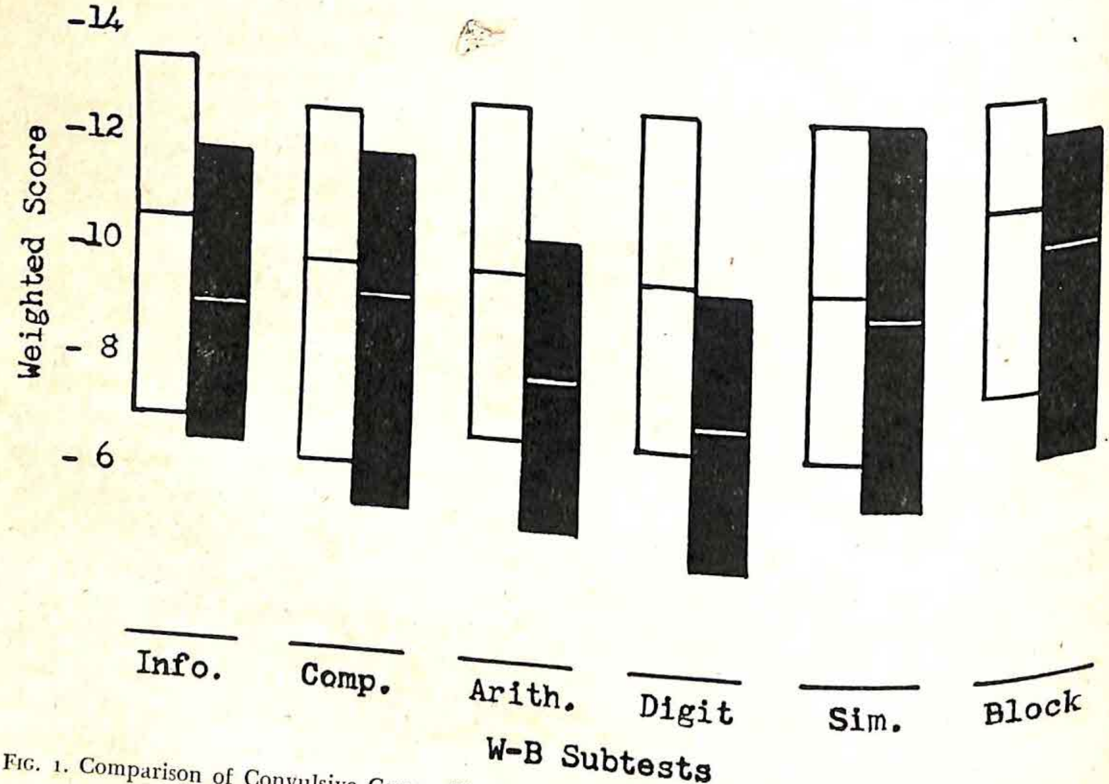


FIG. 1. Comparison of Convulsive Group (Shaded) with Wechsler's Standardization Group of Adults (Open), after Transmutation from Raw to Weighted Scores.

for each of the convulsive patients an index of "loss" according to the procedure proposed by Wechsler. Because, in this study, only certain of [illegible] Wechsler subtests were used, the following formula was adopted for use with equivalent scores on the subtests:

$$\frac{(\Sigma \text{ Information} + \text{Comprehension} \times \frac{5}{3}) - (\Sigma \text{ Arith, Digits, Simil, Block Des} \times \frac{5}{4})}{(\Sigma \text{ Information} + \text{Comprehension} \times \frac{5}{3})}$$

As this index is proposed, the greater the per cent "loss," the greater is the patient presumed to be deficient in those abilities (as measured by Arithmetic, Digits, Block Designs, etc.) which decrease normally with advancing age, and with "organic" deterioration. It is suggested by Wechsler that when the percentage of "loss" is 20 per cent or more, "organic" impairment may be suspected.

Thirty-seven per cent of the convulsive patients show no loss; 46 per cent show loss within the 1 to 20 per cent range which Wechsler regards as inconclusive evidence of "organic" pathology. The remaining 17 per cent of the patients demonstrated loss of 20 per cent or more—indices ranging from 22 per cent to 38 per cent.

Comparison of EEG evaluations for the "loss" group of 10 cases and for the total group of 72 cases shows little difference between the groups.

EEG	Per Cent Total Group	 Loss Group
Abnormal	47	40
Borderline	24	20
Normal	29	40

With no effort toward determination of the statistical significance of the differences between these groups, it seems to be clear that the "loss" patients are no more inclined toward pathological EEG records than convulsive patients in general, who show, to be sure, a relatively high frequency of "abnormal" EEG records.

Discussion. Performance with tasks involved in mental tests, chiefly those of the W-B Scale, indicates that the group of convulsive patients we are studying here does not differ greatly, so far as total IQ is concerned, from the "normal" standardization group. Their performance in certain tasks (mainly Arithmetic and Digits) is significantly less competent compared with their performance otherwise and with the performance of the "control" (standardization) group.

On the basis of his studies of various types of patients at the Menninger Clinic, Rapaport (16) found that patients, whatever their psychopathology, are likely to show in their performance with the W-B verbal tests a deficiency in Arithmetic and Digits particularly. To him, Arithmetic and Digits are "more vulnerable to maladjustment" (16, p. 166). He feels that within the W-B battery of subtests these two tests (in which our convulsives are deficient) are unique in this respect: they have to do with the subject's *relation to reality*. The Digit test, he believes, is a test of *attention*, by which he means "an effortless, passive, unhampered contact with outside reality—a free receptivity" (16, p. 168). Thus the subject who, like many of our convulsives, is low on Digit Span is incapable of this easy, spontaneous relationship to the reality of his environment. In further discussion Rapaport maintains that the ability to attend is in inverse proportion to the presence of anxiety.

Like the Digit test, Arithmetic also measures the patient's contact with reality, according to Rapaport, but "in contrast to attention (measured by the Digit Span test) concentration (which is measured by Arithmetic) may be characterized as an *active* relationship to reality" (16, p. 168). "Attention then may be regarded as automatic, concentration as voluntary[2] effort" (16, p. 169).

If Rapaport's interpretation of the psychological significance of W-B subtests is valid, we might conclude that the relationship to reality of these convulsive patients is poor. It seems likely that anxiety prevents these patients from giving attention—it reduces their receptivity, and this, in turn, reduces their efficiency in handling problems intellectually. They possess the capacity for more effective function than they demonstrate, and this is because their private worries reduce stimulation from the outside, and hence the facility to mobilize abilities for optimal performance.

CONTROL IN THE CONVULSIVE PATIENT

Personality Structure As Revealed in Test Performance

In this phase of the report comparison will be made with data obtained from "normals" to reveal the extent to which the convulsive patients as a group seem to utilize particular kinds of adjustment. Here we shall depend largely on data obtained by means of the MMPI and the Rorschach test.

In Figure 2, mean scores for the MMPI are plotted on a chart designed to compare any group (or patient) with the scores obtained for control subjects of either sex. In this instance the convulsive patients are compared with 293 subjects for *Pa* and *Pt;* 117 for *Mf;* 294 for *Hs, D, Sc,* and *Ma;* and 345 for *Hy*. These designations are explained on page 1. Variability of the patient group is indicated by horizontal lines drawn at plus and minus one standard deviation from the mean.

Figure 2 shows that the convulsive patients demonstrate greater variability in almost all of the main variables of the MMPI; this is most true of *D, Hs, Sc,* and *Hy*. Statistical comparison of the convulsive group with the control group shows (at the .01 level of confidence) increased scores for the former in *D, Mf, Hy, Hs, Pt,* and *Sc;* and (at the .05 level) in *Pd*. Differences for *Pa* and *Ma* are insignificant.

The results indicate that, when compared with the group of controls on which the MMPI was standardized, (*a*) on all main variables, the convulsive patients as a group have mean scores greater than the controls; (*b*) deviation of the convulsive group is *least* for variables *Pd, Pa, Sc,* and *Ma*—variables related in general to *psychotic* nosology; it is *greatest* for *D, Mf, Hy, Hs,* and *Pt*—variables related in general to *neurotic* nosology; (*c*) apparent exceptions to this general conclusion are *Pd (psychopathic deviate), D* (related presumably to *depression),* and *Mf* (tendency to react in ways similar to the opposite sex).

It is our feeling that much of the con-

---

[2] Rapaport adds that "It should be clear that in using the expression 'voluntary' we [refer not to] an independent psychological force, such as in the past has been termed 'will' [but] rather to the energy freely available to the individual for disposal in intellectual processes in contrast to specifically-deployed energies which are not so available" (16, p. 169).

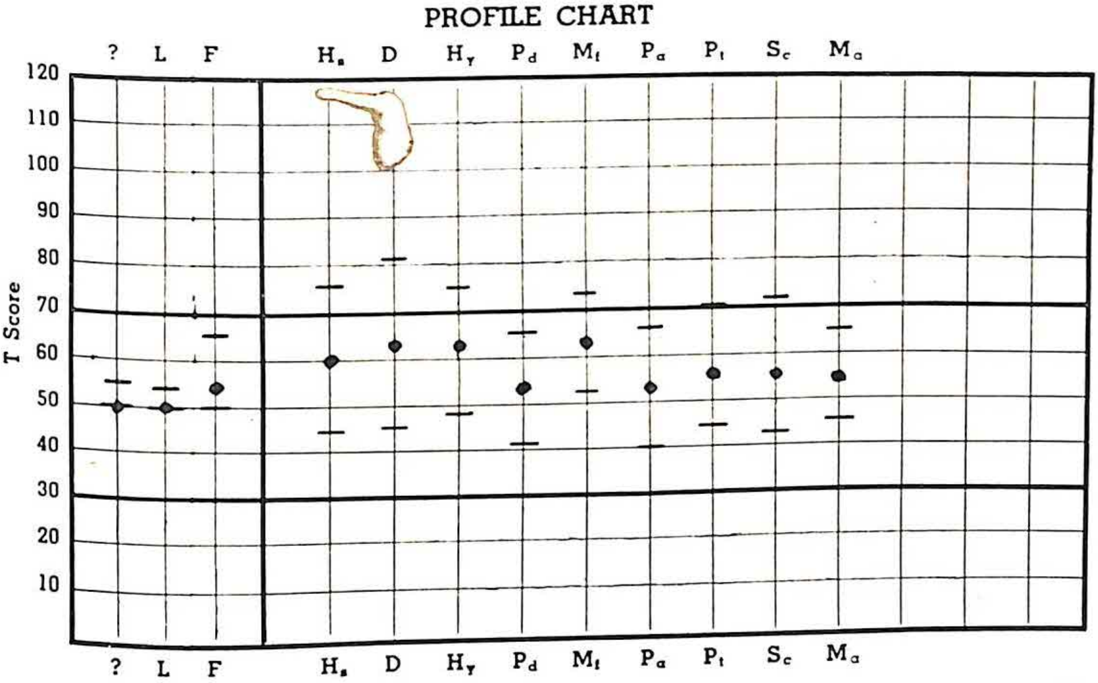


FIG. 2. Comparison of Convulsive Patients (Mean Represented by Dot, ± 1 σ by Horizontals) with the Male Standardization Group, in Terms of *t* Scores.

tent of the *D* scale of the MMPI reflects depression of a mild, nonsuicidal, and nonpsychotic nature, and is therefore neurotic. Regarding *Mf*, a tendency for men to react feminine-wise is consistent with a neurotic (extratensive) adjustment rather than a psychotic (introversive) one. (This will be discussed further in connection with item analysis of the MMPI.) Somewhat the same reasoning is involved regarding *Pd;* this represents a masculine sort of psychopathology, less extreme than *Pa* and *Sc* but in the same direction (away from neurosis, and toward psychosis). Differences between sexes for the standardization groups show that normal control females give higher *D* scores and lower *Pd* scores than control males, a point which bears out the suggestion above.

These results show that in the MMPI situation the convulsive patients behave in a manner less like that of the control men than of the control women. In fact, the mean scores for the convulsive patients (with the exception of *Mf*) are very close on all MMPI variables to the mean scores for normal control women.

A general conclusion from these data regarding scores for the main variables of the MMPI is that the convulsives tend toward psychopathology, particularly of an extratensive, feminine, and neurotic sort. Indeed, the profile for the group of convulsive men is much more like that of normal women than that of normal men, if we exclude the measure of *Mf*, or masculinity-femininity itself. However, the convulsives demonstrate greater variability than the standardization group. The fact that they are less homogeneous than the control group argues against the concept of an "epileptic personality" type.

Data provided by the Rorschach performance provide an interesting check on these results of the MMPI. It is well known that, of the several scores made of Rorschach responses, those for *determinants* are presumed to yield a measure of introversive vs. extratensive tendency, of ego function and therefore of *control.*

In Figure 3 is shown the proportion of total responses which is determined by each of several possibilities, such as movement, color, pure form, etc., for the convulsive group (black columns) as compared with responses (in blocked out columns) for the group studied by Neff and Lidz (14) of 38 men in the middle IQ range. These were soldiers in the Pacific area.

It is seen that the control group excels the convulsive group in *M* (human movement) and *FM* (animal movement). The convulsive group excels the experimental group in *m* (inanimate movement), in *k* (third-dimensional responses), in *F* (pure form), in *C′* (black-or-white as color), and in *FC* (form-color). For determinants such as *FK* (vista), *Fc* (form-texture), and *CF* (color-form), the groups are equally productive. The convulsives show a small increase in *K* (diffusion) and a decrease in *C* (pure color).

According to Rorschach theory, these findings suggest that the convulsive patients are less spontaneous and impulsive —more restrained—than the controls. Both groups *dehumanize* movement. The controls do this by using animal movement (*FM*); the convulsives do so by utilizing movement presumably beyond human control, that is, inanimate movement (*m*). The higher *F* column for the convulsives suggests a greater tendency to sheer repression. The manner of dealing with impulses from within shifts from spontaneity (*M* and *FM*) to *m* and to *k*, a tendency for the convulsive patients to intellectualize and depersonalize. As far as social spontaneity is concerned, the shift from color (notably *C* and *FC*) to shading or achromatic color (*C′*) suggests that depression

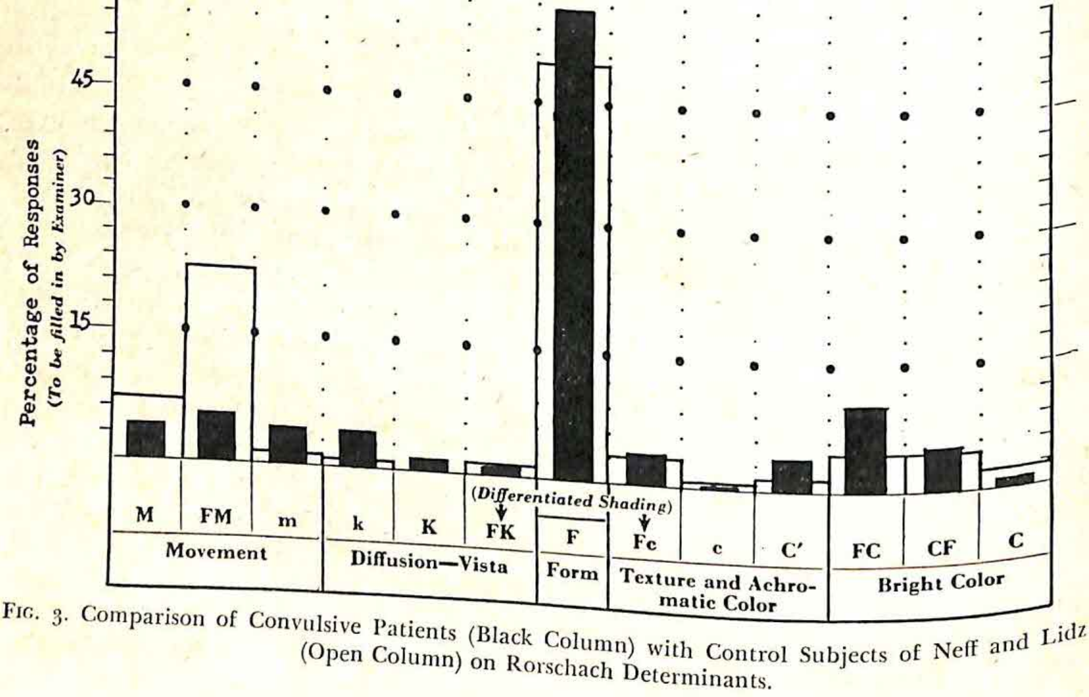


FIG. 3. Comparison of Convulsive Patients (Black Column) with Control Subjects of Neff and Lidz (Open Column) on Rorschach Determinants.

and traumatic experience are important in the history of these patients.

In terms of total productivity, the convulsive group gave a mean of 10.3 responses, in contrast to a mean of 14.2 for the controls. This reflected in great measure a tendency for the convulsives to see nothing, or to reject certain cards. The average number of cards rejected by the convulsive patients was almost two. In order, the most frequently rejected cards, together with the per cent of rejections, were: VII, 32.0; IX, 29.2; VI, 26.4; X, 20.8; III, 19.4; IV, 18.1; II, 16.7; VIII, 12.5; V, 5.6; and I, 1.4.

It is generally recognized that of all Rorschach cards, IX is the most frequently rejected. Of considerable interest is the fact that, among the convulsive patients, Card VII (presumed by many to be the "mother figure") was rejected more frequently than IX, and that it was rejected much more frequently than was IV (the "father figure"). Also of interest is the fact that VI (the "sex" figure) was third in order of rejection. These possibilities are considered in greater detail in the following discussion of *motivation.*

MOTIVATION IN THE CONVULSIVE PATIENT

THE CONCEPT OF SELF

The patient's present situation. It should be obvious that we are dealing here with individuals under a special set of circumstances. In the first place, they are in military service, during a period of intensely active warfare, and with no indication of an end to hostilities. Many of these men had been on active duty for several years. They had been away from home, practically all of them outside the continental limits of the United States. More than half had been in combat areas, and most of these under combat stress.

Secondly, these men (during the period they were studied) were patients in a neuropsychiatric hospital. This means that they were temporarily removed from a situation of imminent orders to combat duty, and in this sense they may have felt protected. For at least a few, as we shall see, hospitalization meant a challenge to demonstrate adequacy or normality.

Certain of the items in the questionnaire administered to the men are of interest in this connection.

Why are you in the hospital? Twenty-six men (36 per cent) write simply "Epilepsy" or "Epileptic." For these men, this single concept seems to account adequately for their present situation. Another 15 men (21 per cent) use the term "epilepsy" or "epileptic" together with some elaboration. Examples are: "Epilepsy (they say)," "Suspicion of epilepsy," "Diagnosis is epilepsy (Jacksonian)," "I'm in the hospital at present because I'm believed to have epilepsy." This second group, who elaborate somewhat on the word "epilepsy," betray at least some doubt that the single diagnostic term will suffice as an accurate statement. One patient writes: "For drinking too much whisky, causing me to pass out; they have me in for epilepsy." Another omits the term epilepsy, but states: "Because I passed out. People who saw me said it was a fit. I don't know, outside of feeling a little groggy afterwards, I feel all right."

Over half the patients, then, bring in the word "epilepsy" in accounting for their hospitalization. Ten more use terms just as simple, such as "blackout spells," "hazy spells," "unconscious spells," "fainting spell," "for passing out," etc. Only a minority discuss their hospitalization subjectively, i.e., in terms of the experience which led to medical attention. Illustrations of these are: "I have fainting spells and don't know what happens," "Because I had a fit while working," "For passing out, reason I don't know."

A conclusion from the characteristic answers to this question about the reason for hospitalization is that these men regard themselves as *having something* (namely, "epilepsy"). They are in the hospital not so much because they need help or need study, but because they have been diagnosed as epileptic. Not all the patients are willing to accept this diagnosis, but most realize that it has been, or may well be given them. In evaluating themselves in this way they use not their own but an outside frame of reference.

A second item in the questionnaire asked *How soon do you think you should go to duty?* followed by *What type of duty do you prefer?*

> Twenty patients (38 per cent) either gave no answer or said "don't know." Ten men (14 per cent) wrote "Never" or some other form of definite negative answer. Nine wanted duty limited to the U.S. ("stateside"). One man wrote the words "As soon as I am" and crossed them out. A quite different group of 16 men (22 per cent) were outspoken in asking for duty immediately: "Any time, repair ship"; "To the battle zone." An extreme illustration of this is M. R., who wrote as his preference, "Hazardous duty in underwater demolition team."
>
> A third group of patients, 14 in all (19 per cent), leave the matter of returning to duty up to the Navy or the hospital: "When the cause of my spells has been determined, flying or aircraft maintenance"; "Whenever the doctor advises, hospital work"; "When the doctor says I am well"; "Up to the doctor, wherever I am put"; "When cured"; "Go any time when doctor thinks so"; "To duty when I am cured of what I got"; "Combat duty, when proven I haven't got it [epilepsy]."

These replies indicate that a group of men hospitalized for convulsive phenomena have a range of attitude about further active duty, with a larger proportion than one might expect expressing preference for active and even combat duty immediately, or "as soon as the doctor says so." There seems to be a high degree of motivation in a sizeable number of these patients to demonstrate that they are not inadequate or incapacitated by "epilepsy," whether they "have" it or not.

Replies to personality questionnaires. One of the test instruments used with these patients was the Cornell Selectee Index. In order to have some basis for comparison, the replies of the convulsive group are presented in Table 1 in terms of the proportion answering "Yes" to each question together with the proportion found characteristic of a group of 325 penitentiary prisoners studied by Wickersham (22). Only those items are presented on which the groups differed with statistical significance at the .01 level of confidence.

Comparing the penitentiary with the convulsive population, we see that the former (as might be expected) have been to reform school, have been arrested, have gotten in trouble because of drinking, drink heavily, and have been in mental hospitals. None of these is true of the convulsives. The prisoners have a less stable work history. The prisoners complain more of constipation and, though the difference is less than 10 per cent, of other somatic difficulties like vomiting blood, stammering, uneasiness, urinating in public, pain in the eyes, and nervous breakdown. More of the prisoners can do good work while others are watching, and more of them have a good appetite.

Naturally the convulsive patients complain more of fits, fainting spells, and dizziness. They complain more of tiring easily, of fatigue on waking, of severe headaches, of pressure or pain in the head. Also, the patients regard themselves much more as worriers. They find it hard to forget an unpleasant experience, and more of them have the feeling of being watched while at work.

These contrasts would suggest that the convulsive patient is more subjective, more neurasthenic in his self-preoccupation than is the prisoner. He is more aware of bodily sensations, such as cold hands or feet, twitching of the

TABLE 1

DIFFERENCES BETWEEN CONVULSIVE GROUP AND GROUP OF 325 PENITENTIARY PRISONERS CERTAIN ITEMS CORNELL SELECTEE INDEX

Questions	Per Cent "Yes"	
	Convulsive Patients	Prisoners
Have you ever had a fit or convulsion?	75	6
Do you frequently get up tired in the morning?	69	45
Do you find it hard to forget an unpleasant experience?	58	40
Do you often shake or tremble?	45	22
Do you get tired easily?	44	30
Do you worry about your health?	44	26
Do you often have spells of dizziness?	44	18
Are you easily upset or irritated?	44	24
Have you been troubled with shyness?	44	26
Do you get spells of exhaustion or fatigue?	43	21
Do you sometimes have nightmares?	41	23
Do you often have pressure or pains in the head?	40	19
Have you at times had a twitching of the face, head, or shoulders?	37	22
Do you frequently feel faint?	33	14
Do you suffer badly from frequent severe headaches?	33	18
Have you had more than five jobs in the last two years?	7	20
Have you been arrested more than three times?	6	65
Were you ever a patient in a mental hospital?	3	15
Has any doctor ever told you you had ulcers of the stomach?	1	6
Do you drink more than two quarts of whiskey a week?	1	22
Have you ever gotten into serious trouble or lost your job because of drinking?	0	26
Do you use dope regularly?	0	8
Were you ever sent to a reform school?	0	38

head, face, or shoulders, trembling, fatigue, and general irritability. Nevertheless, 73 per cent of the convulsives feel well and happy, 81 per cent make friends easily, feel that they are usually treated fairly.

A second test instrument utilizing the questionnaire method, the MMPI, consists of items related not only to somatic complaints, but to a variety of interests and activities as well. Items of the inventory[3] on which the convulsive group answered "True" in significantly greater proportion than the "normal" standardization group ($N=139$)[4] are presented in Table 2 and Table 3. The criterion of significance is a difference at the .01 level of confidence or better.

A glance at the pattern of differences between the convulsive patients as a group and the normal group reveals certain major areas of importance in the psychology of these patients.

a. Understandably, the convulsive patients are much more inclined to admit dizzy and fainting spells and fits and convulsions than are the controls. They report head complaints with much greater frequency.

b. They feel generally less adequate. They are less sure of their memory, their judgment, their ability to concentrate; they have a greater tendency to forget what people say.

c. In greater frequency also, the patients report subjective phenomena such as hot spells, dermal sensations of burning and itching, ears ringing or buzzing.

d. Their sleep is much more troublesome than is that of the controls. This includes more dreams, nightmares, and sleepwalking, as well as insomnia, sensitivity to wakening, etc. Fewer of them feel well rested in the morning.

e. The patients tire more quickly, are less able to work.

---

[3] It is important to consider that only those items of the MMPI were used which contribute toward a score for one of the main variables.

[4] The author wishes to thank Dr. Starke Hathaway for his kindness in submitting data on the standardization group.

TABLE 2

MMPI Items on Which the Proportion of Convulsives Answering "True" is Significantly Greater than that of Control Subjects

Statement	Per Cent Answering "True"	
	Convulsive	Control
Any man who is able and willing to work hard has a good chance of succeeding.	100	89
I would rather win than lose a game.	99	87
I have often had to take orders from someone who did not know as much as I did.	63	42
I have periods of such great restlessness that I cannot sit long in a chair.	60	37
Sometimes I become so excited that I find it hard to get to sleep.	60	33
I frequently notice my hand shakes when I try to do something.	57	27
Parts of my body often have feelings like burning, tingling, crawling, or like "going to sleep."	53	23
I forget right away what people say to me.	32	12
Most of the time I feel blue.	30	5
I cannot keep my mind on one thing.	29	13
Much of the time my head seems to hurt all over.	27	4
I have had periods in which I carried on activities without knowing later what I had been doing.	27	9
Often I feel as if there were a tight band around my head.	23	5
I feel weak all over much of the time.	21	6
I have nightmares every few nights.	21	5
I have the wanderlust and am never happy unless I am roaming or traveling about.	21	6
I find it hard to keep my mind on a task or job.	21	7
Once a week or oftener I become very excited.	20	6
I dislike having people about me.	19	4
At times I have fits of laughing and crying that I cannot control.	13	2
Once a week or oftener I feel suddenly hot all over, without apparent cause.	13	2

f. There is a slight suggestion of confusion, in that more of them report that things sometimes seem unreal, their thoughts race ahead. A small but significantly greater percentage "commonly hear voices without knowing where they come from" (11 per cent vs. 3 per cent).

g. The patients are less inclined to report strong attraction to members of their own sex. They are less fond of recalling "Alice in Wonderland" and playing "Drop the handkerchief." Also, they are less willing to grant sexual freedom to women comparable to that enjoyed by men.

h. On the whole, the patients give a picture of conscientiousness, competitiveness, and absence of humor. More of them worry about their health than do the normals, but fewer have financial and other worries.

Projection of the self in fantasy. Rorschach theory suggests that the individual's concept of himself is best revealed in the *movement* response. Reference to Figure 3 (p. 8) indicates that, including human, animal, and inanimate movement, the convulsive patients use little (40 per cent of them use none). Piotrowski feels that movement (particularly human movement) represents a concern on the part of the patient for the *future*—for future possibilities. Absence of movement in this group of convulsives suggests *caution,* an unwillingness to use the Rorschach situation expressively; it may be inferred that the patients are, as a group, repressed and confused as to their role in life. However, it is interesting and pertinent to examine in some detail the *nature* of movement responses which appear in 60 per cent of the records.

Human movement (the most spontaneous form of all) was used by 22 subjects (30 per cent). Of the total number of human-movement responses, 64 per cent were in response to Card III, which is regarded by all Rorschach workers as most provocative of the human-movement response. These responses are:

two people picking up something
two people working
two persons, kind of bending over
two people, but I don't know they're doing . . . bending down (trying to jitterbug)
couple of people sitting over a fire . . . their hands, trying to keep warm
two persons standing close to a fire, like warming their hands
child's drawing of a couple of (colored) people (leaning)
a design of people doing some kind of a dance
two people pulling something
two men pulling something
couple of people doing a dance together—dancing team
waiters arguing over a spilled tray
two people doing something
two people pulling on something . . . trying to get it away from each other. Could be a man, they both look like women
two men trying to pull something apart
form of two people, fighting, pulling away from each other
two people washing clothes
two people doing *anything*

Of the remaining Rorschach cards, three (Cards I, VI, and VIII) produced no human movement; Card IV produced

TABLE 3

MMPI ITEMS ON WHICH THE PROPORTION OF CONVULSIVES ANSWERING "TRUE" IS SIGNIFICANTLY LESS THAN THAT OF CONTROL SUBJECTS

Statement	Per Cent Answering "True"	
	Convulsive	Control
I like to visit places where I have never been before.	81	95
I seem to be about as capable and smart as most others around me.	80	94
I am about as able to work as I ever was.	73	91
I believe women ought to have as much sexual freedom as men.	70	88
I have little or no trouble with my muscles twitching or jumping.	66	88
My hands have not become clumsy or awkward.	66	86
I hardly ever feel pain in the back of the neck.	64	86
I do not often notice my ears ringing or buzzing.	63	81
My judgment is better than it ever was.	63	87
My daily life is full of things that keep me interested.	61	91
I am always disgusted with the law when a criminal is freed through the arguments of a smart lawyer.	61	83
I believe I am no more nervous than most others.	60	83
I am neither gaining nor losing weight.	60	86
I daydream very little.	59	79
I do not tire quickly.	54	83
I have very few headaches.	53	83
I seldom worry about my health.	53	77
I can read for a long while without tiring my eyes.	49	75
I like to be in a crowd who play jokes on one another.	49	69
I seldom or never have dizzy spells.	44	78
I have never had a fit or convulsion.	43	77
I almost never dream.	36	59
I am against giving money to beggars.	32	55
I worry over money and business.	29	63
I have never had a fainting spell.	26	49
I used to like drop the handkerchief.	24	44
I have never felt better in my life than I do now.	20	50
I drink an unusually large amount of water every day.	19	38
I liked "Alice in Wonderland" by Lewis Carroll.	16	37
In walking I am very careful to step over sidewalk cracks.	11	29
I would like to be a soldier.	11	61
I am strongly attracted by members of my own sex.	10	27
I have a great deal of stomach trouble.	4	18
Everything is turning out just like the prophets of the Bible said it would.	1	9
I think I would like the work of a librarian.	1	9

one ("a man's feet, he's sitting on something"). Five cards produced two each (Cards II, V, VII, IX, and X): Card II (two people playing "pease-porridge-hot" and "design of two people dancing"); Card V ("the hips, legs there, . . . spread out" and "couple of men lying on a hillside"); Card VII ("two old ladies looking at each other" and "a couple of dancers again, comic costume, back to back"); Card IX ("a woman giving birth there" and "the Devil having an argument with himself"); and Card X ("somebody playing pinochle" and "over here in the blue you see a woman bending down, on her hands and knees . . . looks like she's looking at the back of a man's head, who is in a reclining position, and then there's another man, and a baby, lying with his neck stretched out").

A general impression from these human-movement responses is that there is to most of them, again, a quality of *restraint:* people are working, picking up something, bending, warming themselves at a fire, or busy doing anything at all. For a few patients the action is aggressive —people are pulling against one another, pulling something apart, even fighting. By comparison with the human movement characteristic of normal individuals, there seems to be, for these convulsive patients, a higher degree of flexion, suggesting hesitancy and caution.

So far as animal movement is concerned, the convulsive patients, when compared with the control group, reported little. As conventional interpretation would have it, these patients are less infantile in their impulsiveness than are the controls. However, there is among the patients more animal than human movement.

Twenty-five patients (34 per cent) gave animal movement. Of this animal movement, 30 per cent was given for Card VIII. This would be expected in a routine group of Rorschach records, because of the provocation of animals doing something (climbing, eating, crawling, etc.) which this card has in contrast to all the other cards. These are the responses given for Card VIII by the convulsive patients:

bears trying to climb up
climbing animals
wildcat crouched, ready to spring
lizards climbing a tree
rats, or mice, climbing somewhere
two animals climbing up on something
two animals crawling along, going someplace
a couple of bugs crawling on a flower, insects
two animals probably climbing a plant
two animals creeping up
looks like a couple of bears trying to climb somewhere
two animals nibbling at a tree

Here, it is seen, there is considerable emphasis on climbing, and the *attempt* to climb.

Next to Card VIII, Card X elicited the greatest number of animal-movement responses, nine in all. These are:

two animals hiding in a tree
two deer floating . . . jump
two birds flying along here
caterpillars going up a tree or branch
animal, gnawing on trees
animals fighting
an antelope, on his hind legs, "on his getaway"
insects eating on a stalk
animals eating around some stains

Here the emphasis on going up, climbing, or flying is particularly impressive, but there are emphases also on eating and on escape (hiding, "getaway").

Card V evoked seven animal-movement responses, all of which involve flying or flight:

a bird, or such, flying . . . coming at you
a bird, flying
a flying butterfly
moth or butterfly in flight
a bat flies around
looks like a bat, spread out, flying through the air
a bat, in flight

For Card I (which brought forth five animal-movement responses) flight is again in evidence:

a flying animal, a bat
a flying bat
blood-sucking bat, wings spread
crawling bug of some type
a bat, in flight

Cards III and IX provided no animal-movement responses. Of the remaining cards, II gave three ("two bears standing and facing each other"; "two animals, got their noses together, straight up in the air"; and "two butterflies, in flight"). Two were elicited by Card VII ("two dancing dogs" and "two objects, with a butterfly resting its wings on them"); Card IV had one ("a serpent, coming out of a hole"), as did Card VI ("a flying squirrel").

From these animal responses we may conclude that expression is cautious and tentative. There is much flying, crawling, leaping; in other words, effort against the force of gravity which, rather than aggressive, is defensive.

Inanimate-movement responses are expressed by the convulsive patients in most abundance on Card VII. (It will be recalled that this was the achromatic figure most rejected.) There were eight such responses:

clouds, thunderhead . . . boiled up, dark
billowing smoke
looks like a ship burning, or something like that
two Indians looking at each other
smattering of a snowball, against a wall, splashed or smashed
an animal taken apart and just laid all over there
broken clouds, or chain
pair of clouds, closing down at the base

Clouds or smoke, on Card VII, are frequent responses with "normals" but this combination of diffusion with inanimate movement (suggesting free-floating anxiety) is impressive.

Card X was next in order in evoking inanimate-movement (six responses):

two spiders' stepped on
greenish chain, like you pull in older times
two worms fastened together
spiders hanging outside a tree
crabs, hanging on to some kind of animal
rabbit's head, fluid coming out of the eye-sockets

Here again, there is a striking emphasis on things suspended or falling, squashed, or otherwise subject to gravitational force.

Card II, with five inanimate-movement responses, elicits the following:

a mashed bug
an explosion, something burning, flame coming out
a crawdad, smashed out
clowns, tied together
something disfigured, mutilated

Here, for Card II, the feeling of catastrophe suffered—already accomplished—is very strong.

There were four responses utilizing inanimate movement for Card IV.

a fountain (spraying)
dark clouds, separating before thunder
animal, his eyes popping out
tree, branches bowing over

There were four also for Card IX:

something squashed
water has eaten away the land
an oil fire . . . something burning . . . a lot of heat . . .
looks like a waterfall

The remaining movement responses of an inanimate sort were: three for Card III ("a couple of animals hanging"; "black ink smeared"; and "two men, locked together") and for Card VIII ("two animals stretched out"; "a flare from an explosion"; and "a forest fire, in technicolor"); two for Card I ("a squashed beetle" and "a plane, burning up") and for Card VI ("an altar, lighted up . . . sparks coming out" and "Christ! looks like somebody took a man down,

. . . laid him open . . . his buttocks out, and left his penis hang").

Parental identification and discrimination. Within recent years, considerable attention has been given by some investigators to the differential response to individual Rorschach cards. In particular, Card IV, considered to be the masculine or "father" figure, has been compared with Card VII, the "mother" figure.

In the discussion of *control* it was shown that Card VII (the mother figure) is the most rejected card while Card IV (the father figure) is rejected much less frequently. On the basis of this finding alone, one might conclude that these convulsive subjects are more inhibited by the figure of the mother, or the feminine, than by the masculine.

We have seen that inanimate movement responses (*m*) were given most frequently for Card VII, the mother figure. Since *m* responses are regarded as revealing anxiety, we might infer that anxieties regarding the feminine role and one's relation to it are prominent in these convulsive patients. Those concerning the masculine role and figure reveal less anxiety.

A somewhat more detailed analysis of responses of these convulsive patients to Card IV in contrast to Card VII is provided by the following tabulation of the *sorts* of responses given, in terms of proportion:

PERCENTAGE OF TOTAL RESPONSES

Response	Card IV ($N=89$)	Card VII ($N=91$)
Rejection	13.5	26.4
Human Movement (*M*)	1.1	4.4
Animal Movement (*FM*)	2.2	3.3
Inanimate Movement (*m*)	4.5	11.0
Third Dimension (*k*)	4.5	13.2
Diffusion (*K*)	1.1	11.0
Pure form (*F*)	48.3	15.4
Texture (*c* and *Fc*)	14.6	5.5
Shading (*C'*)	10.1	9.9

It appears from this tabulation that convulsive patients reject Card VII more frequently than they do Card IV but they also react much more actively to Card VII in other ways: they give more movement—human (*M*), animal (*FM*), and particularly *m* (inanimate movement). They project more in the nature of third dimension (*k*) and diffusion (*K*). Pure form is much less for VII than it is for IV. Other than pure form responses, IV excels only in *c* and *Fc* (the texture responses). This suggests that these patients are, in relation to the father figure, not only repressed but *cautious and tactful* by comparison with their relation to the mother or female figure. The female figure seems to elicit an active participation, either of flat denial or of *effort,* to intellectualize or otherwise rationalize. The male figure, on the other hand, is handled in a matter-of-fact way, with caution and with kid gloves, but without excitement.

Of the few patients (twelve) who rejected the father figure, half also rejected the mother figure, and the remaining four gave *K* or *k* responses (maps, clouds, etc.).

In handling the father figure (excepting the 6 who rejected it) the 24 patients who rejected the mother figure were fairly factual, realistic, and spontaneous.

We might infer that it is fairly characteristic of the convulsive patient to reject the feminine while preserving a fairly well-integrated pattern of (impersonal) relationship with the masculine. When this pattern of masculine relationship is distorted, as it is for a few of the subjects, the feminine relationship seems also to be tenuous and anxious, but reflected in intellectualization as a method of defense, probably of long duration. The feminine image seems to force on these patients projection, introspection, objectivity, detachment, and remoteness.

The self in the sexual situation. The convulsive patients gave few overt sex responses. It was noticed, however, that they frequently rejected Card VI, the "sex figure." A fourth of the patients did this, and responses for a sizable group of the remainder were characterized by features of repression and anxiety of which *F* (49 per cent) and the frequency of *C′* (black and white shading) sort of response were the chief indicators. The *C′* response was the most frequent determinant other than pure form (*F*). Characteristic responses of this sort are:

a monument, a goddess, lighting up, sparks coming out
a camouflaged gun—fur
a bear skin, sir—dark
a skinned hide—the color of it
a skin (because of color and shape)
a man laid open, buttocks out, penis hanging (and) a hide, shape, and color
a pelt—shape and color
a hide—shape and color
an X-ray—shape and color
black print of a lady—lungs

These responses are good form—in fact, the hide or pelt is a popular response. But the frequency with which the color of the figure partly determines the percept is high.

One might infer that, in the sexual situation, these convulsive patients are formal and repressive; the situation is dangerous, and they are socially ill at ease and depressed.

The Thematic Apperception productions. For purposes of tapping motivational factors in the personality of these convulsive patients, the Thematic Apperception Test was used. The stories told by the patients were scored by means of an adaptation of the technique recommended by Aron (2).

In general, the stories were brief and sterile; the typical patient was hesitant to give free rein to his imagination. Much of the content was simple description, with dynamics or action quite secondary.

For five of the pictures, stories were so stereotyped and homogeneous as to be of little use in the present analysis. Nonproductive pictures were, as Stein (20) describes them:

> *Picture 3.* On the floor against a couch is the huddled form of a boy with his head bowed in his right arm. Beside him on the floor is a revolver (20, p. 3).
> *Picture 9.* Four men in overalls are lying on the grass taking it easy (20, p. 7).
> *Picture 12.* A young man is lying on a couch with his eyes closed. Leaning over him is the gaunt form of an elderly man, his hand stretched out above the face of the reclining figure (20, p. 8).
> *Picture 14.* The silhouette of a man (or woman) against a bright window. The rest of the picture is totally black (20, p. 9).
> *Picture 17.* A naked man is clinging to a rope. He is in the act of climbing up or down (20, p. 11).

The remaining five pictures were considerably effective in eliciting material suitable for analysis. These were pictures 4, 6, 7, 13, and 18.

In order that the reader may have at least a rough norm for comparison, Stein's description of what is usually seen by the young adult male patient will be given for each picture as it is discussed here in connection with the performance of convulsive patients.

Picture 4. ("A woman is clutching the shoulders of a man whose face and body are averted as if he were trying to pull away from her.") According to Stein:

> The stories to this picture are usually concerned with a conflict situation between the couple in the foreground. The man may desire to leave the woman because he has a plan which he wishes to put into action but the woman wants to keep him with her. The conflict may be stated in general terms as an "argument" or the story may involve the "eternal triangle" with the seminude figure in the background as the man's mistress or girl friend. These stories frequently reflect the patient's difficulties in marital ad-

> justment or his attitudes toward women and sex (20, p. 4).

It was characteristic of the convulsive patient not only to be sensitive to the conflict between the man and the woman in the picture, but to be aware of it intensely. He revealed himself as belonging to one of three groups, depending on which sex was finally dominant in the story: the man or the woman. It happened that of 71 patients, 27 belonged to the male dominant group and 27 to the female dominant group. Seventeen were ambivalent. Typical stores are:

> (Male dominant)
>
> Well, the man is about to do something that the girl doesn't like, apparently. . . . She's trying to keep him from going ahead and doing it. He might be mad at somebody . . . going to beat him up. It resembles a movie. . . . She's trying to restrain him. From the look in his eyes, it looks as if he's got a mind of his own.
>
> (Female dominant)
>
> This might have been an affair of a couple, and he's evidently gonna shove off on her or leave her, and she's evidently in anxiety that he's gonna leave and she doesn't want him to leave. From the expression . . . (Will she be successful?) I think she will. There's not quite enough expression of determination on his face.
>
> (Ambivalent)
>
> Well, I can see that the man is mad, and the woman is trying to soften him up. . . . Whether I can tell a story, I don't know. . . . Arguments can start over almost anything. (Will she be successful?) Well, I imagine she will, . . . they generally are . . . although I don't know. . . . He's kinda wild-eyed, stubborn. . . .
>
> (Second illustration) Well, let's see. . . . It looks like this man has been out with another woman. I presume this is his wife here, wanting him to come back home. He's undecided. I don't know what the outcome will be.

In the light of the clear tendency for the convulsive patient to reveal his identification with the masculine, with the feminine, or both, it becomes interesting to examine any differential in response to subsequent pictures.

Picture 6. ("A short elderly woman stands with her back turned to a tall young man. The latter is looking downward with a perplexed expression.") According to Stein the scene is structured in this way by young adult males:

> The elderly woman and young man in this picture are usually seen as mother and son. The son may be seen as asking his mother's permission to do something that he had planned for a long time. He may want to leave home to undertake a job in another city, to get married, or to enlist in the armed forces. His desires usually conflict with his mother's. In most stories we are also told how the conflict is resolved. These stories reflect the patient's attitude toward his mother and the factors which may have caused friction at home (20, p. 5).

Among the convulsive patients in their stories for Picture 6, aggression by the son toward the mother was expressed by 56 per cent of the male-dominant group, 37 per cent of the female-dominant group, and 29 per cent of the ambivalent group. This is a story characteristic of aggression toward the mother:

> That looks like a mother and son. Evidently he's done something that his mother is disappointed in him. . . . She's acting like she doesn't want to talk to him now, but I think in the end—like a mother always does—she'll take him back.

However, mood qualities of tension or of dejection were prominent in the ambivalent group (29 per cent) and in 11 per cent of the male-dominant; this was absent in 96 per cent of the female-dominant group. In the female-dominant group it was characteristic (in comparison with the other groups) for the mother to have suffered at the hands of someone else (not the son). Also, there was a tendency for the son to be subject to aggression or dominance by the mother, quite consistent with the sort of story told for Picture 4. This dominance of mother over son also characterized many of the stories of the ambivalent group.

Picture 7. ("A gray-haired man is looking at a younger man who is sullenly staring into space.") By co[illegible]rison with the mother-son picture (No. 6[illegible] the stories told for Picture 7 were in striking contrast. Stein describes, as typical of the young adult male in the service, the following as characteristic of Picture 7:

> The younger man is frequently seen as having come to the older man for advice or the two men are discussing a problem of mutual interest. These stories frequently indicate the patient's attitude toward adult males and also some of the problems about which he may be concerned. At times the stories to this picture reflect the patient's attitude toward therapy (20, p. 5).

Among the convulsive patients, there is almost no aggression on the part of the son; for all three groups (male- or female-dominant, and ambivalent) the theme is one of dominance by the father and of affiliation with the father by the son. There is, in general, an absence of the tension and dejection which characterize the stories told for the mother-son figure. Here is a typical story for Picture 7.

> Well, it looks like it may be a father advising his son—something his son is very interested in—taking it all in.

Picture 13. ("A young man is standing with downcast head buried in his arm. Behind him is the figure of a woman lying in bed.") For the normal male in the service, according to Stein, Picture 13 is treated as follows:

> Sexual plots are most frequent in the stories told to this picture. The young man is usually seen as contemplating or having had sexual intercourse with the woman on the bed. The woman may be the young man's wife, a girl friend, or a prostitute. These stories usually reveal the patient's attitude towards women and sex.
>
> Another frequent theme is that the woman lying on the bed is the young man's wife, who is either sick or dead. The young man's feelings are then described. This theme usually indicates the patient's hostile attitude towards his wife or women in general (20, p. 9).

In their stories for Picture 13, the group of patients who (for Picture 4) are female-dominant reveal least need sex: 15 per cent vs. 21 per cent for both the male-dominant and ambivalent groups. Aggression toward the woman is prominent in the female-dominant group: 44 per cent vs. 30 per cent for the male-dominant, and 15 per cent for the ambivalent group.

Illness and death to the woman resulting from actions (other than those of the hero) characterize 47 per cent of the stories of ambivalent patients, 30 per cent of those female-dominant, and 20 per cent of those male-dominant.

From the manner of dealing with Pictures 4, 6, 7, and 13, one obtains a strong impression that the convulsive patient is extremely sensitive to the possibility of conflict between the sexes. In fact, sexuality seems to be a matter of struggle—of battle—in which one or the other sex is victor. Catastrophe to the woman is a pervading theme, more characteristic of patients uncertain about sexual dominance or sensitive to female supremacy. We have, thus, some strong confirmation of the frustration-aggression hypothesis advanced by Dollard, Doob, Miller, Mowrer, and Sears (4).

Picture 18. ("A man is clutched from behind by three hands. The figures of his antagonists are invisible.") Responses to Picture 18 were of particular interest because we are dealing with the phenomenon of convulsion. The description of the picture and of stories it evokes in young adult males, according to Stein, are as follows:

> Patients are frequently concerned with their inability to develop an adequate rationale for the three hands that appear in the picture. In the common stories the man in the picture is seen as either drunk or having been in an accident and the hands belong to the people who are helping him. Or the hero is being attacked from the rear and the hands

> belong to his attackers. These stories reveal the patient's anxieties concerning aggression that may be directed against him or the patient's attitude toward alcoholism or drug addiction (20, p. 11).

Within the groups of convulsive patients, there were no significant differences between subgroups divided (according to response to Picture 4) as male- or female-dominant, or ambivalent. The great majority of the stories described the central figure as experiencing "press" from outside, rather than needs within. Of these, press Aggression was most frequent (39 per cent). (The hero suffers attack by another—he is hit on the head, beaten, or robbed.) Next in order was press Nurturance. (The hero is being helped, or carried, or otherwise cared for.) This characterized 27 per cent of the stories. A third sort of press, experienced by 17 per cent, was Affliction; the hero is ill or intoxicated.

SUMMARY

In this study a group of 72 convulsive patients has been compared wherever possible with "control" subjects. This comparison has suggested that, according to the data of this study, convulsive patients differ among themselves as much as they differ from the controls (or the population generally). This is argument against the concept that the convulsive personality is unique, or that there is, indeed, a convulsive personality.

For a few variables, the convulsive group demonstrates restricted variability. These are the test performances Digits and Arithmetic on the Wechsler-Bellevue Scale of Intelligence, and the scales for *Ma* (Hypomanic tendency) and *Mf* (tendency toward interests of the opposite sex) on the Minnesota Multiphasic Personality Inventory. This relative homogeneity on the part of the convulsive patients for Digits and Arithmetic accompanies a lowered mean on these tests, something which, according to Rapaport ([illegible]), is characteristic of many patients examined at the Menninger Clinic. From this it is inferred that the convulsives, so far as mental test performance is concerned, are much like neuropsychiatric patients in general.

The mean difference of the convulsive group from the controls for MMPI *Ma* is insignificant statistically, so that this warrants no further comment. Mean deviation from the norm for *Mf* is as great, however, as that for any other variable of the MMPI. This indicates that, so far as interests are concerned, the convulsives are like the opposite sex and quite uniformly so. This is consistent with the fact that their mean profile for the MMPI is much more like that of the control females than like that of the control males.

Intellectually, the group of convulsives (with a mean IQ of 98) does not differ from the general population. However, 5 per cent perform like mental deficients; 16 per cent (including one of the "mental deficients") perform in a manner which, according to the criteria suggested by Wechsler (21), is indicative of "loss" due presumably to brain damage. (What proportion of the control or "normal" population demonstrate such loss is not shown in Wechsler's data.)

In their intellectual performances, particularly with Digits and Arithmetic, the convulsives reveal anxiety, which presumably serves to reduce in general their effectiveness in the mental test situation.

So far as ego structure, or control, is concerned, there is strong indication that, at the level of conscious operation the convulsive patients as a group are *extratensive;* they deviate from the normals more on the "neurotic" variables of the MMPI than on the "psychotic" variables; they deviate in the direction of feminine interests and personality structure as this may be indicated by the MMPI profile. Extratension is shown furthermore by the ratio of human movement to sum of color on the Rorschach, with color predominating over movement.

In looking for explanation of these phenomena in terms of dynamics, the psychological data suggest that defense against anxiety must be recognized at both the extratensive and the introversive ends of the continuum. Extratensively, these patients show immaturity in the utilization of social techniques such as tact and diplomacy (as might be suggested, for example, by texture responses on the Rorschach); they tend to preoccupation with somatic difficulties (as indicated by responses to the MMPI and particularly the Cornell Selectee Index) and feelings of personal inadequacy, verging toward neurotic depression. They are lonely people, who lack sophistication and *savoir faire,* and rely on illness and inability to justify their isolation.

On the introversive side, however, there is evidence also for a rationalizing, intellectualizing sort of defense. Many of these patients were remarkably quick to size up their current life situation as a matter of having been diagnosed by someone else as "epileptic." Rational accounting of this sort was utilized in the Rorschach situation by a tendency to "maps" and other intellective responses. It was demonstrated more conclusively by the utilization of inanimate movement instead of spontaneous human movement or the primitive, impulsive animal movement. These patients are sensitive to strong forces within themselves, but they are unwilling to accept them as their own. In the face of these promptings, they are moved, therefore, to depersonalize them, to see them as outside influences, abstract and symbolic.

Seeking further—to ask what lies behind this adjustment which seems compounded of self-doubt, self-denial, and uncertainty—the psychological data give certain motivational clues. The Rorschach protocols suggest strongly that the situations of particular threat to these convulsive patients are sexual in nature; these men tend to reject sexuality outright, to repress all spontaneity. Because complete repression is usually impossible, there breaks through an acceptance of sexuality which is realistic, but colored with a spirit of inadequacy and sadness.

Their responses to masculine and feminine figures of the Rorschach suggest further that these convulsive patients have acquired a fairly realistic and functional integration with masculine possibilities; in their man-to-man relationships they are passive, but cautious and tactful and generally at ease. Their relationship to feminine possibilities is in sharp contrast. In the heterosexual situation, the initial reaction of these convulsives is repression and active rejection. The aggression with which this rejection takes place is revealed in the active manner with which they deal with the feminine figure, through *projection.* With the feminine figure, in an atmosphere of intellectualizing and abstracting and externalizing, movement comes into the foreground. But it is not human movement which dominates the picture; it is depersonalized, inanimate movement. In the heterosexual situation these patients are not only anxious to quell their anxiety, but they feel called upon to take *action* of some sort.

These impressions of threat in the heterosexual situation are reinforced by the stories told for the TAT pictures. Here it is characteristic that man-to-man situations are fairly relaxed, objective, and impersonal. The hero is usually passive. When women are involved, the hero is aroused, emotional, and under pressure to *do* something. While the number of patients who revealed dominance and the number who revealed submission in the heterosexual situation are equal (with a sizable group ambivalent), it was the submissive group of patients who revealed the greater aggression toward the woman, and the lesser degree of simple need sex.

The convulsive patients showed, in their TAT responses as well as in the Rorschach protocols, a high degree of sensitivity to conflict between the sexes—a battle in which one sex or the other is the victor. Very often the final result is catastrophic to the woman, though not always because of the man's behavior.

These findings suggest that the psychodynamics of *identification* with the mature masculine role is somehow distorted or off balance for many of these patients. One might ask as to their early history in terms of experience with parents. About a third of them lost at least one of their real parents before the age of sixteen, the majority of these the father (through death in nine cases and by desertion, separation, or divorce in five more). In addition to this obvious evidence of disruption to the family unit, there is evidence in many more of the histories of parental conflict, and of "nervousness" or "spells" on the part of parents, particularly the mother.

The fantasy of these patients as it is expressed in the Rorschach, and particularly the TAT, suggests strongly that they are confused about sexual roles, that they have failed in many instances to integrate within their personalities a masculine identification that is maturely effective in heterosexual relationships. They seem frustrated by the female, but they are hesitant also to express their aggression toward her directly. They attempt to dissipate this aggression, first by repression (which usually does not work), then by rationalization. Here again they are unable to achieve a satisfactory outlet, and it is in this setting that illness, attributed by the patient to forces outside himself and beyond his control, emerges in the form of "fits," which finally become completely justified for many of them by the diagnosis of "epilepsy," even though it is characteristic for the patient to want to disguise this because of its implications vocationally and socially.

BIBLIOGRAPHY

1. Arluck, E. W. A study of some personality characteristics of epileptics. *[illegible]h. Psychol., N.Y.*, 1941, **263**. Pp. 77.
2. Aron, Betty. *A manual for analysis of the Thematic Apperception Test: A method and technique for personality appraisal.* Berkeley, Calif.: Willis E. Berg, 1949.
3. Beck, S. J., Robin, A. F., Thiesen, W. G., Molish, H., & Thetford, W. M. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1951, **30**, 241-298.
4. Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.
5. Guirdham, A. The Rorschach test in epileptics. *J. ment. Sci.*, 1935, **81**, 870-893.
6. Harrower-Erickson, Molly R. Personality changes accompanying cerebral lesions. II. Rorschach studies of patients with focal epilepsy. *Arch. Neurol. Psychiat., Chicago,* 1940, **43**, 1081-1107.
7. Harrower-Erickson, Molly R. Psychological studies of patients with epileptic seizures. In W. Penfield and T. C. Erickson, *Epilepsy and cerebral localization.* Springfield, Ill.; Chas. C Thomas, 1941. Pp. 546-574.
8. Hathaway, S. R., & McKinley, J. C. A multiphasic personality schedule: I, construction of the schedule. *J. Psychol.*, 1940, **10**, 249-254.
9. Jessico, C. M., Findley, K. H., & Richards, T. W. Observations in a mild form of epilepsy. *Minn. Med.*, 1947, **30**, 1949-51.
10. Kettleman, Natalie. A Rorschach approach to the dynamics of epileptic behavior. Unpublished Master's thesis, Northwestern Univer., 1949.
11. Klopfer, B., & Kelley, D. M. *The Rorschach technique: A manual for a projective method of personality diagnosis.* Yonkers-on-Hudson, N. Y.: World Book Co., 1942.
12. Lisansky, Edith S. Convulsive disorder and personality. *J. abnorm. soc. Psychol.*, 1948, **43**, 29-37.
13. Morgan, Christiana, & Murray, H. A. A method for investigating fantasies: The Thematic Apperception Test. *Arch. Neurol. Psychiat., Chicago,* 1935, **34**, 289-306.
14. Neff, W. S., & Lidz, T. Rorschach pattern of normal subjects of graded intelligence. *J. proj. Tech.*, 1951, **15**, 45-57.
15. Piotrowski, Z. A. On the Rorschach method and its application in organic disturbances of the central nervous system. *Rorschach Res. Exch.*, 1936-37, **1**, 23-40.
16. Rapaport, D. Diagnostic psychological testing: The theory, statistical evaluation, and diagnostic application of a battery of tests. Chicago: Year Book Publishers, 1945. Vol. I.
17. Richards, T. W. Modern clinical psychology. New York: McGraw-Hill, 1946.
18. Rorschach, H. *Psychodiagnostics.* (Trans. by P. Lemkau and B. Kronenberg.) New York: Grune & Stratton, 1942.
19. Stauder, K. *Konstitution und Wesenänderung der Epileptiker.* Leipzig: *George Thieine,* 1938 (Review by Louise Collins in *Rorschach Res. Exch.*, 1944, **8**, 38-41).
20. Stein, M. I. *The Thematic Apperception Test: An introductory manual for its clinical use with adult males.* Cambridge, Mass.: Addison-Wesley Press, Inc., 1948.
21. Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams and Wilkins, 1944.
22. Wickersham, F. M. A factor analysis of the responses of psychopathic personalities on the Cornell Selectee Index. Unpublished Ph.D. dissertation, Northwestern Univer., 1949.

Accepted for publication February 25, 1952.

PREFACE

In the years immediately following World War II, when there was a good deal of speculation and little dependable information about the adjustment of veteran students to college, the idea of making a thorough investigation of the problem was suggested by the Carnegie Corporation. After considerable thought and discussion of the possible values of such a study and the methods of obtaining and analyzing data, the investigation described here was undertaken by the College Entrance Examination Board with the support of the Carnegie Corporation. Following the merger which led to the formation of the Educational Testing Service, the study was carried to completion by ETS.

A great many people contributed to the st [illegible] riety of ways. It is unfortunate that acknowledgment by name cannot b [illegible] hose people at the various colleges and universities who permitted t [illegible] ade and who supplied the data; these people cannot be named because [illegible] ot to reveal the identities of the participating institutions.

Acknowledgment is due to Dr. John Do [illegible] e Carnegie Corporation, and to Dr. O. C. Carmichael, Preside [illegible] Foundation for the Advancement of Teaching, for their helpf [illegible] nning the general objectives and outlines of the study. Mr. He [illegible] lent of the Educational Testing Service, contributed many use [illegible] criticized portions of the manuscript. The technical advice of Dr. [illegible] er, Professor Harold Gulliksen, and particularly Professor S. S. Wilks [illegible] valuable. Mrs. Judith Aronson was in charge of computing and Mr. Harry Garrison supervised the punching and tabulating operations. Mr. Donald Peterson assisted in making arrangements with the participating colleges for obtaining data. The manuscript was read by Dr. William W. Turnbull, who made many helpful suggestions.

Norman Frederiksen
W. B. Schrader

Vol. 66, No. 15 Whole No. 347, 1952

Psychological Monographs: General and Applied

The Academic Achievement of Veteran and Nonveteran Students[1]

Norman Frederiksen and W. B. Schrader
Educational Testing Service, Princeton, N.J.

I. PLAN OF THE STUDY

A. Purposes of the Study

In the period 1945-1950, American colleges were crowded with students, a large proportion of whom were veterans of World War II. In part, the influx was an outcome of a new feature in the educational scene—the educational provisions of the federal law commonly called the GI Bill of Rights. Many of these veteran students would not have attended college without the aid thus provided. At the beginning, considerable concern was expressed regarding the possible effects of combat and of other features of wartime service upon the adjustment of veterans to the life of colleges typically designed for a less widely experienced student body.

The experience of university faculties quickly demonstrated that the more pessimistic views were unfounded. University teachers and deans reported that the veterans were alert and industrious students, that their influence on the undergraduate body was wholesome, and that the incidence of events that could be traced to battle shock was much smaller than had been anticipated. Numerous [illegible] published in newspapers [illegible] magazines in which rather [illegible] of the scholastic success [illegible]nts were presented. Vet- [illegible] almost universally reported [illegible] or to nonveterans in aca[illegible]nt.

[illegible]ndation for the Adva[illegible]ing became interest[illegible] of the academic success [illegible]udent in college, and the College Entrance Examination Board was requested to make a study investigating the relationship between veteran-nonveteran status and academic success in college. The College Board agreed to conduct such an investigation, with the support of the Carnegie Corporation. The primary objectives of the study were to answer the following questions:

1. Do veteran students make better grades in college, in relation to their ability, than nonveteran students?

2. What light does information about background, attitudes, and other qualities throw upon veteran-nonveteran differences?

3. How do veterans who could not have attended college without the financial assistance provided through the GI Bill compare with veteran students who were financially able to attend college?

---

[1] This monograph is based on Chapters I, II, and III of *Adjustment to College: A Study of 10,000 Veteran and Nonveteran Students in Sixteen American Colleges* (privately printed in a limited edition by the Educational Testing Service, 1951). The study was conducted by the College Entrance Examination Board and the Educational Testing Service with the support of the Carnegie Corporation of New York. That Corporation does not, however, necessarily endorse the statements made or views expressed herein.

This report is concerned with the first of the three questions—the comparison of veteran and nonveteran students with respect to grades relative to ability.

The analysis of the data needed to study this question would furnish information relating to numerous other questions of interest to educators. How well can scholastic success in college be predicted from aptitude-test scores and measures of high-school achievement? What improvement in prediction can be effected by combining various predictors of college success? Are there differences between veterans and nonveterans with regard to the relative effectiveness of predictors for forecasting college achievement? Of what significance i[illegible] ential in time of taking apt[illegible] relation to time of enteri[illegible] veterans who enter coll[illegible] en handicapped during th[illegible] in comparison with non[illegible] at a warm-up or refresh[illegible] be desirable? In the [illegible] n whose education w[illegible] ne war, is a period of [illegible] necessary, or does he con[illegible] with enthusiasm which leads to temporary overachievement? Questions of this sort were taken into account in designing the study.

The best opportunity for making such a study presented itself in the fall of 1946, which was the time when the maximum number of veterans (particularly veterans who had actually experienced combat overseas) were enrolling in college. It was felt that the academic year 1946-1947 was the optimal time to study the question of the effect of war service on college achievement. Students enrolled at that time would possess to a marked degree the characteristics which make veterans different from nonveterans, and would be sufficiently numerous to yield statistically stable results.

B. Selection of Colleges

It was desired that the study of veterans' achievement in college be based on data obtained from a variety of types of institutions, including coeducational and men's colleges, private colleges and colleges supported by state or municipal funds, large and small institutions, colleges with various curricular emphases (such as liberal arts, engineering, and agriculture), both heavily endowed colleges and colleges with less endowment, and colleges representing various geographical regions of the United States. It was also desired that pairs of colleges with roughly the same characteristics be chosen. Further restrictions in the choice of colleges included adequate numbers of students, availability of suitable data, and willingness of the college to participate in the study.

A tentative list of 36 colleges which seemed to meet the criteria for inclusion in the study was prepared. Letters were sent to the presidents of the 36 colleges by Dr. O. C. Carmichael of the Carnegie Foundation for the Advancement of Teaching; each letter briefly described the objectives of the study. A few days later a letter and a brief questionnaire were sent from the College Entrance Examination Board to the same persons. Anonymity of the colleges in the published report was promised to the institutions who wished to participate in the study.

The questionnaire sent to the college presidents was designed to provide information which would be useful in judging whether or not the institution was an appropriate one to include in the study from the standpoint of adequacy of data. It furnished a convenient and uniform method for the college to report: (*a*) the number of veteran and nonveteran students entering each college or division of the university in the fall of 1945 and in the fall of 1946; (*b*) information regarding the program of intelligence or scholastic-aptitude testing, including names of tests used, whether tests were normally taken near the completion of secondary school or at the time of entrance to college, and the proportion of students having scores; (*c*) whether or not some over-all measure of achievement (such as average grade or grade-point average) was routinely available for these students; (*d*) whether specific courses were taken in common by all or by a large proportion of the freshman students; and (*e*) whether or not some measure of scholastic success in high school was available for a large proportion of the entering freshmen. A question was included to ascertain whether or not the institution wished to participate in the study.

Twenty-six of the colleges indicated that they wished to participate in the veterans' study and returned questionnaires. Seven institutions could not or did not wish to participate, and from the remaining three colleges there was no reply or a noncommittal reply, with no reply to follow-up letters. Nine of the 26 colleges which expressed willingness to participate in the study were not included because of an inadequate number of cases in one or more of the groups to be studied or because of the lack of certain crucial information. Data were obtained from 17 colleges. One institution was dropped from the study after the data had been collected, since needed data were available for too few students, so that the statistical analysis is based on groups from 16 colleges and universities.

The participating colleges will be referred to in this report by code names. The private colleges were given pseudonyms which are common American surnames assigned at random with no attempt to make the name carry any implication as to any characteristic of the institution. The publicly supported institutions, including both state and municipal colleges, were given geographical names which describe in a general way the location of the institution. The following is a list, by geographical section, giving the code name and a brief description of each institution:

Eastern City University: A coeducational publicly supported college of arts and science located in an eastern city.

Adams University: A private university for men located in an eastern city.

Stewart University: A private university for men located in a small eastern city.

Miller University: A private coeducational university located in an eastern city.

Midwest State University: A large coeducational state university located in a midwestern city.

Midwest City University: A coeducational, publicly supported university located in a midwestern city.

Littletown State University: A coeducational state university located in a small midwestern city.

Harris University: A private coeducational university located in a small midwestern city.

Evans University: A private coeducational university located in a midwestern city.

Central State University: A large coeducational state university located in a small midwestern city.

Tur[illegible] *University:* A private coeducational university located in a midwestern city.

Mi[illegible] *Technological University:* A coeducational [illegible] college emphasizing agriculture, [illegible] other applied arts, located in a small [illegible] city.

Middle S[illegible]*ity:* A large coeducational state university located in a midwestern city.

Douglas University: A private coeducational university located in a southern city.

Sou[illegible] *Technological University:* A coeducational land-grant college, emphasizing agriculture, engineering, and other applied arts, located in a southern city.

Western State University: A coeducational state university located in a western city.

C. Collection of Data

A member of the College Board staff visited each of the participating institutions, with the exception of Western State University, for the purpose of making detailed arrangements for the collection of data.[2] Prior to the visits, a letter was written requesting that preliminary arrangements be made for selecting a supervisor to have charge of the data collection.[3]

---

[2] Mr. Donald Peterson and the senior author of this report were responsible for making these arrangements with the cooperating colleges.

[3] It is regretted that individual acknowledg-

The original plan of the study called for the administration of achievement tests in mathematics, physics, or chemistry, where appropriate, for the purpose of supplementing regular course grades by objective examinations which could be used as additional criteria of academic success. Most colleges, however, were unable to cooperate in this phase of the study. Arrangements for administering achievement tests were made at only one institution. However, when the tests were actually administered and scored, it was found that the difficulty of the tests was not appropriate for the students; the results were therefore not used in this study.

1. *Groups for Whom Data Were Obtained*

Three main types of [illegible] were planned: (*a*) comparisons [illegible] veteran students who entered in [illegible] of 1946 with male nonveterans [illegible] entered at the same time; (*b*) similar studies for students who entered in the fall of 1945; and (*c*) comparisons of male veterans whose college careers were interrupted by war service and who returned to college after the war with the best available control group.

Collection of data on adequate numbers of students who entered in the fall of 1946 proved to be relatively straightforward; similar efforts to obtain groups of adequate size among students entering in 1945 proved to be difficult because the nonveteran groups were frequently inadequate in size. As a result, only one of the twenty-five groups actually studied was limited to students entering during the academic year 1945-1946. In a few other instances, students entering in the fall of 1945 were pooled with students entering in the fall of 1946 for analysis.

In the collection of data for the study of veterans whose schooling was interrupted, the basic plan was to limit the group to veterans who had completed one year prior to interruption and one year after interruption. For three groups, this plan was followed. In two additional groups of interrupted veterans, however, it was found desirable to increase the flexibility of the defining pattern in order to obtain larger groups.

Logically, it would seem desirable to use male nonveterans who had experienced a similar interruption for comparison with the interrupted veterans. In practice, however, it proved to be impossible to locate more than thirty interrupted nonveterans in any one group. Consequently, nonveterans whose schooling was not interrupted, but who were otherwise as similar as possible to the interrupted veterans, were selected for this purpose. Thus, for three of the groups, male nonveterans who entered in the fall of 1945 and who completed two full academic years consecutively formed the comparison group. For the two interrupted groups including veterans having a flexible pattern of interruption, the uninterrupted nonveteran comparison group was composed of students entering in the fall of 1939. Care was taken, of course, to ensure that there was no reason to doubt the comparability of grading standards before determining that the control group was suitable.

2. *Criterion Data*

The criterion data included average grades, as determined by the college in which the student was enrolled or by calculation from data supplied by the college. In addition, grades in specific courses, such as English or mathematics,

---

ment in this report is precluded by the decision that complete anonymity of the colleges should be maintained.

were obtained in cases where the course had been taken by all or practically all the students.

Other types of criterion data which were considered and included in certain instances were such variables as academic standing (e.g., "good standing, probation, dropped"), number of visits to the health department for psychiatric problems, and attendance. Because of the very small proportion of students in these extreme categories, however, and for other practical reasons, these data were not used in the statistical analysis.

3. *Predictor Variables*

The colleges had originally been selected in part on the basis of the availability of adequate predictor data. Scores on some intelligence or scholastic aptitude test were available in all colleges; the most usual test of this sort was the American Council Psychological Examination (ACPE). In most cases, some measure of high-school achievement was also obtained, such as high-school average grade or rank in high-school class. In addition to the intelligence-test score and high-school grade, other data which were thought to be potentially good predictors of college achievement were included, achievement tests used by the college for admissions or for sectioning being the most common example. In some colleges composite scores or predicted freshman grades were available, and these were also included. Still other variables of theoretical interest were obtained, although they were not expected to be good predictors of achievement. Examples of such variables are date of high-school graduation and date of taking aptitude and achievement tests; there was considerable variability on these factors for veteran students in certain institutions.

A variety of methods of recording data were used—hand-copying to rosters, photostats, microfilm, and punched cards—depending upon such considerations as cost and the facilities available at the particular institution.

D. Analysis of Data

The primary purpose of the analysis of academic data was to determine whether or not veteran students made higher grades, in relation to their ability, than did nonveteran students. In order to obtain as precise a comparison as possible, it was considered essential to analyze data separately for each university, and, in institutions of complex organization, separately for each college or division. Analyses were based on groups homogeneous with respect to institution attended, the division within that university [illegible] which enrolled, and time of entrance to college. Each such group included both veteran and nonveteran male students. A total of twenty-five separate groups were defined by this process. The groups within each institution are shown in Table 1.

Colleges where differentiation into various divisions is delayed until after the freshman or sophomore year are classified as colleges of arts and science in this study. In all, 17 groups were from colleges of arts and science (so defined), 4 were from engineering colleges, 2 from colleges of business, 1 from an agricultural college, and 6 groups involved interrupted veterans from arts and science and technological colleges.

1. *Preparation of the Academic Data*

In defining the specific students from each college to be included in the statistical analyses and in making the numerous detailed decisions required before analysis could begin, primary consideration was given to two objectives: (*a*) to

TABLE 1

THE TWENTY-FIVE GROUPS I[illegible]LUDED IN THE STUDY, ARRANGED ACCORDI[illegible]O UNIVERSITY

University	Division	Time of Entrance	Number of Cases MV*	MN*
Adams	Arts & Science	Fall 1946	531	694
Adams	Arts & Science	Fall 1945†	134	111
Central State	Arts & Science	Fall 1946	466	166
Central State	Arts & Science	Fall or Winter 1945	135	59
Douglas	Arts & Science	Fall 1946	77	119
Eastern City	Arts & Science	Fall 1946	53	147
Eastern City	Arts & Science	Fall 1945†	70	99
Evans	Arts & Science	Fall 1946	283	94
Harris	Arts & Science	Fall 1946	105	146
Littletown State	Arts & Science	Fall 1946	103	107
Littletown State	Business	Fall 1946	142	65
Middle State	Engineering	Fall 1946	352	98
Midwest City	Arts & Science	Fall 1946	83	72
Midwest City	Engineering	Fall 1946	167	171
Midwest State	Business	Fall 1946	232	58
Midwest Technological	Agriculture	Fall 1946	140	102
Midwest Technological	Agriculture	Fall 1939†	57	106
Midwest Technological	Engineering	Fall 1946	271	128
Midwest Technological	Engineering	Fall 1939†	140	215
Miller	Arts & Science	Fall 1946	425	193
Southern Technological	Engineering	Fall 1945 or Fall 1946	120	50
Stewart	Arts & Science	Fall 1946	187	348
Stewart	[illegible] Science	Fall 1945†	55	70
Taylor	[illegible]cience	Fall 1945 or Fall 1946	100	101
Western State	[illegible] Science	Fall 1946	433	222

* The abbreviations MV and [illegible]nd for male veteran and male nonveteran respectively.

† The date given is for nonveterans. The veterans with whom they are compared entered college at various times and had their college careers interrupted by war service.

bring into sharp focus any basic difference in relative achievement between veteran and nonveteran students; and (*b*) to provide data which would be as comparable as possible for the 25 groups chosen for analysis.

For practical reasons, the detailed planning of the study and the preparation of the academic data for analysis developed concurrently. In order that all analyses be based on the same students within a particular college group, the inclusion of any variable on which data were not complete required the exclusion from the analysis of students who lacked data on that variable. Thus in the final determination of the members of each of the 25 subgroups it was necessary to balance the information to be obtained from a particular variable against the proportion of cases to be dropped and the size and representativeness of the sample remaining, if the variable were included. Fortunately, two facts aided greatly in making the decisions: first, many of the scores most appropriate for the study were available for substantially all cases; and second, it was necessary or desirable to exclude students lacking data on certain variables (e.g., first-year grades) as a means of defining the sample.

2. *Generalized Description of the Samples*

The students included in the various studies were male veteran students or male nonveteran students who belonged to one of the 25 groups chosen for study. These groups were defined in terms of the college or university, the college class (e.g., freshman in the fall of 1946) and the university division (e.g., college of arts and science). In general, each student included had earned a definite amount of college credit at specified periods of time. He had complete data on all variables chosen for statistical analysis in his group. He had not attended, as a civilian student, any other college or university; and in most groups, he had not received substantial credit (10 quarter hours) in specific courses for college work he had taken while in the armed services. In universities having several divisions, he had been enrolled in the designated division during the entire defined period, which was typically one academic year. Except for the editing on credit for armed-service college-training programs, the procedures were uniform for veterans and nonveterans. Typically, the veterans and nonveterans were not separated until the editing was completed.

Direct comparisons of grades of veterans and nonveterans (excepting only the "interrupted" groups) are thus limited to grades earned at the same time, in the same university, and in the same division within the university by veteran and nonveteran students. Students who have done part of their academic work at another college are excluded and veteran students who had extensive college training in basic academic subjects during their military career are usually excluded from the comparisons.

3. *The Method of Analysis*

The first step in the actual analysis was to compute the intercorrelations of predictor and criterion variables.[4] For each group, separate tables of intercorrelations were computed for male veterans and male nonveterans. For certain institutions (Midwest Tech, Middle State, Western State, Douglas, and Adams) a larger number of variables was studied than for the remaining institutions. The more complete analyses were in general made in instances where they were justified by the size of the sample or by the availability of variables thought to be of particular interest.

The basic design of this study required a comparison of the achievement of veterans and nonveterans after allowing for the effects of ability differences between the two groups. Since it was decided that the appropriate procedure for allowing for ability differences involved the use of regression equations, a more general approach to the comparison was followed. In effect, the question was raised: would the same regression line (or plane) provide an adequate description of achievement in relation to ability for both groups? It is apparent that there are several ways in which the regression equations might differ. Thus, for a given set of predictors, grades of one group might be predicted more accurately than would be true of the other group. Again, the relative weights for obtaining best prediction might differ in the two groups. Finally, the constant terms in the two regression equations might be different. Each of these possible differences is meaningful. For the purposes of this study, however, results are most readily interpreted if it turns out that the predictors yield equally effective prediction (as measured by the standard error of estimate) and that the predictors may be assumed to have the same slopes in both groups. If these two conditions hold, the difference in constant terms of the regression equation provides a general measure of the difference between the

---

[4] The extensive tabulations and analyses required by this study were executed by the Department of Statistical Analysis, of which Dr. Ledyard R Tucker is director. Particular acknowledgments are also due to Mr. Harry Garrison, who was in charge of the International Business Machines work, to Miss Henrietta Gallagher, who coordinated the analysis, and to Mrs. Judith Aronson, Head of the Computing Section.

the per cent of *veterans* excelling the average *nonveteran* is always reported. Percentages of less than 50 then indicate superiority of the nonveteran subgroup. Percentages greater than 50 indicate superiority of the veteran subgroup.

The criterion which was uniformly used in the analyses of covariance was freshman average grade, point-hour ratio, or some similar index based on course grades obtained during the freshman year. Grades in specific courses were also used in a limited number of analysis of covariance studies. The predictors varied from school to school, but typically two measures, combined through use of multiple correlation techniques, were employed: a measure of high-school achievement, such as average grade or rank in class, and a test or composite of test scores. The most commonly used test was the Ame[illegible] Council Psychological Examinati[illegible]cores on College Entrance Examination Board tests were used for two colleges, and in a number of instances a composite score based on various scholastic-aptitude and achievement tests or tests of tool skills was employed.

A slightly different approach was used [illegible]he case of interrupted veterans. Here it w[illegible]s assumed that the veteran's freshman [illegible]rades, earned before war service, furnished the best possible predictor of grades earned after discharge from military service. In the typical study of interrupted veterans, therefore, the regressions of second-semester sophomore grades on first-term freshman grades for interrupted veterans and uninterrupted nonveterans were compared. The semester or quarter average grades falling between the first term of the freshman year and the last term of the sophomore year were not used in the analysis. The term just prior to induction, it was thought, might have suffered because of knowledge of the impending induction into the armed forces; and the term occurring immediately after the return from service might not have been typical, either because of need to readjust to academic life or because of a temporary enthusiasm leading to overachievement.

II. RESULTS OF THE ANALYSIS

In this section the results of the analyses of the academic data are presented. In discussing the results of comparisons in the various groups studied, the following order will be followed: (*a*) students of arts and science in private universities; (*b*) students of arts and science in state and municipal universities; (*c*) students of engineering in state and municipal universities; (*d*) students of agriculture and of business in state and municipal universities; and (*e*) interrupted veterans and their uninterrupted nonveteran control groups. Within each of these main divisions, those groups upon which the more extensive analyses were done will generally be considered first.

A. GRADES OF VETERAN AND NONVETERAN STUDENTS IN ARTS AND SCIENCE COLLEGES OF PRIVATE UNIVERSITIES

1. *Adams University*

Adams University is a large private university for men. All candidates for admission are required to present scores on College Entrance Examination Board tests, along with evidence regarding their secondary-school achievement. For the 1946 group, about two-thirds of the entering students had attended private schools. Although specific courses are not required of freshmen, the selection of courses is limited, and all freshmen carry a uniform number of courses. The group of students included in this study may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed a full year's work during the academic year 1946-1947.

The predictors used in the comparison of veterans and nonveterans at Adams were the Verbal and Mathematical scores of the College Entrance Examination Board Scholastic Aptitude Test (SAT). The criterion measure is First-Year College Average Grade. Grades are "percentage" grades based on a 100-point scale. Additional predictors studied were the Adjusted School Rank, and a Predicted Grade, computed by the university, which is based on adjusted rank in school and College Board test scores. The adjustment of the rank in school is based on past records at Adams of students from each particular secondary school. The intercorrelations of the variables are shown in Table 2.

"Date of Tests" was also studied in order to determine the relationship of amount of time elapsing between testing and college entrance to measures of achievement in college. Many of the veterans had taken the College Board test prior to war service, while practically all nonveterans took the tests in the spring just prior to entrance at Adams University. The mean date of testing for veterans was found to be between 1944 and 1945, with a standard deviation of 1.12 years. It is apparent that there is sufficient variation among the veterans to make the results of the analysis meaningful.

The correlations involving year of testing are all low; they range from —.08 to .21. The correlations with term averages are .03 and .06. The highest correlation (.21) is with the Verbal score of the Scholastic Aptitude Test. This positive relationship might be accounted for in terms of higher standards for admission in the more recent years; however, such an hypothesis is not borne out by the correlations with the SAT-Mathematical score and the Adjusted School Rank, which are .02 and .08 respectively. The slightly higher correlation with the SAT-Verbal score may merely reflect growth in vocabulary and other verbal abilities with age. By and large, it appears that date of taking the aptitude tests is a matter of little importance within the limits of age and time found in this study. This finding is of considerable significance in the interpretation of data in this investigation, since in most groups the veteran students were tested after war service, while nonveterans were tested soon after graduation from high school.

The correlation of First-Semester Average Grade with Second-Semester Average Grade is .81 for each of the subgroups. This value may be considered as an indication of the reliability of semester grades, although it is presumably an underestimate. The reliability of average grade for the freshman year may be estimated at .90 or higher, using the Spearman-Brown formula.

The veteran subgroup is seen to have obtained lower mean scores than the nonveteran subgroup on all the variables on which a comparison is possible. It is particularly interesting to note that the veterans were predicted to earn a mean average grade of about 72 and they actually earned a First-Semester Average Grade of 75.4, while the nonveterans were predicted to have a mean average grade of 74.8 and actually earned

TABLE 2

Intercorrelations of Scholastic Aptitude Test Scores, Date Tests Were Taken (for Veterans Only), Adjusted School Rank, Predicted Grade, and First-Year College Average Grades

Adams University, College of Arts and Science, Freshmen, 1946–1947

		Male Nonveterans (N=694)							
		SAT-V	SAT-M	Date of Tests	Adjusted School Rank	Predicted Grade	1st-Year Average	Mean	S. D.
Male Veterans (N=531)	SAT-Verbal		.17		.31	.65	.39	565	86
	SAT-Mathematical	.23			.33	.44	.30	594	80
	Date of Tests	.21	-.02						
	Adjusted School Rank	.29	.37	-.08		.82	.60	76.2	7.0
	Predicted Grade	.62	.46	.16	.76		.65	74.8	6.0
	First-Year Average	.41	.33		.55	.63		75.8	6.4
	Mean	536	576	44.4	72.2	72.0	75.1		
	Standard Deviation	93	84	1.12	7.7	6.9	6.1		

a First-Semester Average of 75.7, a somewhat smaller difference.

It was found that the best single predictor, for both veterans and nonveterans, is the Predicted Grade, and that the Adjusted School Rank alone was nearly as good, especially for nonveterans. The Predicted Grade was tried as a variable for use in controlling ability in the analysis of covariance. It was found, however, when this variable was used, that the slopes of the regression lines were significantly different (that for nonveterans being steeper), so that the interpretation of the results of later steps would be doubtful. It was also found that the Adjusted School Rank was responsible for the different slopes. The greater slope of the regression of grades on Adjusted School Rank for nonveterans means, of course, that for a given increment in school rank there was on the average a greater increase in average grade for nonveterans than for veterans. One can only speculate as to the reason for this finding; but perhaps it is related to the nonintellectual factors which influenced high-school achievement and which have somehow been modified in veterans during the years of military service. Since the College Board test scores, used alone, did not have significantly different slopes for veterans and nonveterans, they were chosen as the predictors to be held constant. It must be added that this choice resulted in some loss of predictive effectiveness.

Results pertaining to the analysis of covariance are shown in detail in Table 3. In the first section of the table (I) are shown the intercorrelations, means, and standard deviations for the variables selected. The criterion is First-Year College Average Grade. The correlations in this instance are seen to be very similar for veterans and nonveterans. The nonveterans obtained higher average scores for both the Verbal and Mathematical parts of the Scholastic Aptitude Test, and their Freshman Average Grade was also slightly higher. Variability of test scores was slightly greater for veterans, as shown by larger standard deviations, but the nonveterans were slightly more variable with respect to average grades. Part II of Table 3 shows that the SAT multiple correlations, based on the two SAT scores, are .48 and .46 for veterans and nonveterans respectively. For both subgroups combined, the multiple correlation is .47. In interpreting these correlations, it is important to remember that the tests were used in selection of the students; the resulting curtailment in vari-

TABLE 3

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Adams University, College of A[illegible] and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: SAT-Verbal	SAT-Math.	1st-Yr. Average Grade	Mean	*SD*	*N*
1. Scholastic Aptitude Test-Verbal	MV MN		.23 .17	.41 .39	536 565	93 86	531 694
2. Scholastic Aptitude Test-Mathematical	MV MN	.23 .17		.33 .30	576 594	84 80	531 694
3. First-Year College Average Grade	MV MN	.41 .39	.33 .30		75.1 75.8	6.1 6.4	531 694

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .48; for MN, .46; and for the combined group, .47.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	2.008	1	Between .10 and .20
B. Equality of slopes	0.733	2	Between .50 and .70
C. Equality of intercepts	1.213	[illegible]	Between .20 and .30

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.36
Advantage expressed in standard error of estimate units	0.06
Per cent of veterans excelling the average nonveteran	52
Level of significance of difference (from IIIC above)	Not significant

ability of test scores has presumably lowered the correlations to some extent.

Part III of the table presents the results of the three significance tests in the analysis of covariance. Differences in errors of estimate and slopes are not significant. The intercepts of the regression planes, also, are not significantly different; when ability (as measured by SAT scores) is made equivalent, the grades earned by veterans and nonveterans are too similar to warrant the conclusion that either group is superior.

Part IV of the table highlights the findings: The veterans have a slight advantage over nonveterans of equivalent ability; the advantage, expressed in Adams University grade units, is .36. (In terms of the mean *grades* shown in Part I of the table, the *nonveterans* have an advantage of .7.) The advantage of the veterans, expressed in standard error of estimate units, is only .06. Perhaps the most meaningful measure of the veteran students' advantage is the percentage of veterans found to exceed the average nonveteran in grades adjusted for ability differences; in the case of Adams, this turns out to be only 52. If there were no difference, this percentage would of course be 50. The difference, as noted above, is too small to be statistically significant.

2. *Stewart University*

For the Stewart University students entering as freshmen in 1946, the analysis design was much like that for Adams. At Stewart, intercorrelations among College Board Scholastic Aptitude Test scores, Adjusted School Rank, date of taking admissions tests, and First-Year College Average Grades were obtained. In comparing the relative achievement of veteran and nonveteran male students, allowance was made for three predictors: Scholastic Aptitude Test-Verbal, Scholastic Aptitude Test-Mathematical, and

Adjusted School Rank.

Stewart is a private institution for men, similar in many respects to Adams University. Students are required to present College Board scores, along with their secondary-school records, for admission. As at Adams, the majority (about three-fourths) of the 1946 entering students had attended private schools. As freshmen, Stewart students select their program within a restricted framework. The group of entering freshmen may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed a full year's work during the academic year 1946-1947. Since all students carry a fixed number of courses, it was not necessary to consider variations in academic load in defining the sample.

The statistical results involved in the comparison of achievement of veteran and nonveteran students are presented in Table 4. The predictors again included the College Board SAT scores, both Verbal and Mathematical. Their predictive value was very nearly the same as at Adams. Adjusted School Rank was also used at Stewart, and its predictive value was high (r = .53 for veterans and .62 for nonveterans). The multiple correlations (based on all three predictors) were higher than for Adams because of the inclusion of the measure of high-school achievement. As at Adams, the nonveteran students on the average had higher scores on all the predictors and also a higher First-Year Average Grade. The actual difference in mean grades is .31 on the grading system used at Stewart. Stewart uses a seven-step grading system; in this analysis, the highest grade was given the value 7, the next 6, and so on, 1 being the lowest.

The analysis of covariance results show that the hypotheses of equal errors of estimate and equal slopes of regression planes are not dis-

TABLE 4

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Stewart University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Subgroup	Correlation with: SAT-Verbal	SAT-Math.	Adj. School Rank	1st-Yr. Average Grade	Mean	*SD*	*N*
1. Scholastic Aptitude Test-Verbal	MV		.21	.29	.40	536	86	187
	MN		.09	.31	.40	570	82	348
2. Scholastic Aptitude Test-Mathematical	MV	.21		.32	.28	570	77	187
	MN	.09		.37	.25	590	74	348
3. Adjusted School Rank	MV	.29	.32		.53	4.12	.76	187
	MN	.31	.37		.62	4.70	.78	348
4. First-Year College Average Grade	MV	.40	.28	.53		4.76	.77	187
	MN	.40	.25	.62		5.07	.83	348

II. Multiple Correlations (Variables 1, 2, and 3 vs. Variable 4): for MV, .60; for MN, .66; and for the combined group, .65.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.014	1	Between .90 and .95
B. Equality of slopes	3.375	3	Between .30 and .50
C. Equality of intercepts	2.143	1	Between .10 and .20

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.09
Advantage expressed in standard error of estimate units	0.14
Per cent of veterans excelling the average nonveteran	56
Level of significance of difference (from IIIC above)	Not significant

TABLE 5

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Douglas University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	H. S. Average Grade	1st-Yr. Average Grade	Mean	*SD*	*N*
1. ACPE Total (raw score)	MV		.13	.35	106	19	77
	MN		.32	.55	117	20	119
2. High-School Average Grade	MV	.13		.52	3.26	2.04	77
	MN	.32		.65	4.07	2.30	119
3. First-Year College Average Grade	MV	.35	.52		1.14	.64	77
	MN	.55	.65		1.40	.75	119

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .59; for MN, .74; and for the combined group, .71.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.0506	1	Between .80 and .90
B. Equality of slopes	2.048	2	Between .30 and .50
C. Equality of intercepts	0.0587	[illegible]	Between .80 and .90

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.02
Advantage expressed in standard error of estimate units	0.04
Per cent of veterans excelling the average nonveteran	.52
Level of significance of difference (from IIIC above)	Not significant

proved. The difference between veteran and nonveteran students in intercepts of their regression planes is well within the range of chance expectancy. The advantage of the veteran is again found not to be significant when account is taken of differences in ability.

In addition to the variables directly involved in the analysis of covariance, attention was given to the year during which College Board tests were taken. The results for this variable were similar to those found at Adams. While practically all the nonveteran students were tested in 1946, the mean date of testing for veterans was midway between 1944 and 1945, with a standard deviation of one year. The correlations of year of testing with other variables ranged from .10 (with Adjusted School Rank) to —.04 (with SAT-M and First-Year Average Grade). The correlation with SAT-V was .07 at Stewart. The time of testing was again found to be a factor of negligible importance within the conditions of this study.

3. *Douglas University*

Douglas is the first of many groups for which American Council Psychological Examination scores were available. It is the only liberal arts group for which the relation of year of high-school graduation to the other variables was studied.

Douglas University is a coeducational private university located in a southern city. Admission to Douglas depends primarily upon the student's secondary-school record. In the 1946 group, about 45 per cent had attended private schools. Certain required courses are set up for freshmen in arts and science, required subjects being English, mathematics, and social science. The group studied at Douglas met the following requirements: all entered as beginning freshmen in the fall of 1946 and completed three full quarters during the academic year 1946-1947.

The predictors used in the comparison of veterans and nonveterans at Douglas were total score on the ACE Psychological Examination and High-School Average Grade. The high-school grade, which was expressed in letters, was converted into numerical form as follows: A =8, A— = 7, B+ = 6, B = 5, B— = 4, etc. The criterion measure is First-Year College Average Grade. Grades were based on a four-category scale, the units of which had numerical values of 3 to 0.

In addition to the variables included in the analysis of covariance, year of high-school graduation was studied for male veterans only. The average veteran in this group was graduated from high school in 1944; there was considerable variability as indicated by a standard deviation of 1.7 years. The correlation of year of graduation with ACPE total score was —.06; with High-School Average Grade, it was —.18; and with First-Year College Average Grade, it was —.28. Thus there was a slight tendency for the veterans who had been out of school longer to have higher grades both in high school and college.

The results of the analysis of covariance are shown in Table 5. The American Council Psychological Examination total score and High-School Average Grade were chosen as predictors. The multiple correlations were found to be .59 and .[illegible] for veterans and nonveterans respectively, and .7[illegible] for both subgroups combined. The higher validities for nonveterans is undoubtedly due in part to their greater variability.

Differences in errors of estimate and slopes of the regression planes were found to be no greater than would be expected by chance, and the difference in intercepts of the regression planes was also not significant. When account is taken of differences between the subgroups in ability, the mean difference in freshman average grade is only .02 (in favor of the veterans) as compared with a difference of .26 (in favor of the nonveterans) when no adjustment is made. In standard error of estimate units, the difference is .04; 52 per cent of the veterans excel the average nonveteran.

TABLE 6

Comparison of Average Grades Earned by Veteran and Nonveteran Male Students

Harris University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: Composite Test	H. S. Rank	1st-Yr. Average Grade	Mean	*SD*	*N*
1. Composite Test	MV		.46	.55	164	28	105
	MN		.54	.52	175	32	146
2. High-School Rank (converted score)	MV	.46		.49	15.0	3.0	105
	MN	.54		.65	17.2	3.4	146
3. First-Year College Average Grade	MV	.55	.49		1.20	.72	105
	MN	.52	.65		1.43	.71	146

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .61; for MN, .68; and for the combined group, .65.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.901	1	Between .30 and .50
B. Equality of slopes	4.024	2	Between .10 and .20
C. Equality of intercepts	0.676	1	Between .30 and .50

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.06
Advantage expressed in standard error of estimate units	0.11
Per cent of veterans excelling the average nonveteran	54
Level of significance of difference (from IIIC above)	Not significant

4. *Harris University*

The analysis of data for students in Harris University was limited to that necessary for a comparison of relative achievement of veteran and nonveteran students after allowing for differences in high-school standing and a composite score based on five entrance tests.

Harris is a coeducational private midwestern college of arts and science which is bound by ties of tradition to one of the major religious denominations. Harris students are drawn from the upper half of their high-school graduating classes. Freshmen choose their own program within a series of general requirements, only English being specifically required. For study at Harris University, the following group of students was selected: students who entered as beginning freshmen in the fall of 1946 and who completed substantially a full year's work during the academic year 1946-1947.

The predictors used at Harris included High-School Rank and a Composite Test score. (High-School Rank was converted to a standard-score scale having a mean of 13 and a standard deviation of 4.) The composite score, which is computed routinely at Harris, is based on the American Council Psychological Examination (1942 edition) and four Form T Cooperative tests: the Cooperative English Test and the three Cooperative General Achievement Tests in social studies, natural sciences, and mathematics. The composite score was obtained by adding to the ACPE total score twice the sum of the scaled scores on the four achievement tests, after both the ACPE total and the sum of the scaled scores had been converted to standard scores based on local norms. The criterion as usual is the First-Year College Average Grade. A six-step grading system is used at Harris; in this analysis, the numerical equivalents of the steps were 3, 2, 1, 0, −1, and −2.

The results of the analysis are shown in Table 6. The nonveterans at Harris proved to be

TABLE 7

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Miller University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with [illegible]			Mean	*SD*	*N*
		ACPE	H. S. Rank	1st-Yr. Average Grade			
1. ACPE (1946) Total (raw score)	MV		.26	.41	111	20	425
	MN		.32	.39	112	20	193
2. High-School Rank (converted score)	MV	.26		.43	14.8	3.3	425
	MN	.32		.58	15.8	3.4	193
3. First-Year College Average Grade	MV	.41	.43		1.49	.56	425
	MN	.39	.58		1.50	.64	193

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .53; for MN, .62; and for the combined group, .56.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	1.136	1	Between .20 and .30
B. Equality of slopes	7.027	2	Between .02 and .05
C. Equality of intercepts	2.732	1	Between .20 and .30

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.07
Advantage expressed in standard error of estimate units	0.14
Per cent of veterans excelling the average nonveteran	56
Level of significance of difference (from IIIC above)	Not significant

TABLE 8

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Evans University, College of Arts an[illegible] Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with		Mean	*SD*	*N*
		ACPE	1st-Yr. Average Grade			
ACPE Total (converted score)	MV		.51	12.0	3.3	283
	MN		.45	13.1	3.1	94
First-Year College Average Grade	MV	.51		2.10	.61	283
	MN	.45		2.05	.60	94
Validity coefficient for combined group:			.49			

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.0285	1	Between .80 and .90
B. Equality of slopes	0.148	1	Between .70 and .80
C. Equality of intercepts	5.504	1	Between .01 and .02

III. Difference between Subgroups w[illegible]h Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.15
Advantage expressed in standard error of estimate units	0.28
Per cent of veterans excelling the average nonveteran	61
Level of significance of difference (from IIC above)	2%

superior on the average to veterans on both predictor measures and also with respect to freshman average grade. Validity coefficients are satisfactorily high. The greater validity of High-School Rank for nonveterans may be due in part to the greater variability of nonveterans on this measure.

The analysis of covariance results show that the hypotheses of equality of errors of estimate and equality of slopes of regression planes are not disproved, and that the difference between the intercepts of the generalized regression planes is not significant. The veteran subgroup is slightly superior. Fifty-four per cent of the veterans exceed the average nonveteran with respect to freshman average grade at Harris. The advantage of veterans is clearly not significant.

5. *Miller University*

As at Harris, the analysis of the academic data at Miller was limited to that required for comparing the relative achievement of male veteran and nonveteran students. The predictors employed were the American Council Psychological Examination total score and High-School Rank. The criterion was First-Year College Average Grade.

Miller is a large coeducational private university located in an eastern city. Students in the College of Arts and Science are typically drawn from the upper three-fifths of graduating classes of accredited high schools. The only course required of all freshmen is English; as usual, however, the formulation of the student's program is determined in part by broader requirements. The group selected for study may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed a full year of work (eleven or more hours each semester) during the academic year 1946-1947.

Results of the analysis of data for Miller students are shown in Table 7. In interpreting means and standard deviations, the following information may be useful: The grading system at Miller is based on a five-step scale; the numerical equivalents of the categories range from 3 through 0 to −1. For use in statistical analysis, High-School Rank was converted to a standard scale having a mean of 13 and a standard deviation of 4.

Male veterans tend to excel nonveteran students of equivalent ability. Fifty-six per cent of the veteran subgroup exceed the average nonveteran; but the difference is not significant. The difference in slopes of the regression planes, however, casts some doubt on the accuracy of the evaluation of the difference in intercepts.

6. *Evans University*

For students entering Evans University as freshmen, the comparison of achievement between veterans and nonveterans was carried out using only a single predictor—total score on the American Council Psychological Examination—as a basis for equating aptitude. The criterion, as usual, was First-Year College Average Grade. Evans is a coeducational, private, church-connected college of arts and science located in a midwestern city. Admission is based typically upon the high-school record of the student. At Evans, English is a required subject; the remaining requirements in the freshman program allow some choice of subjects. The group chosen for analysis at Evans was limited to students entering as beginning freshmen in the fall of 1946 who completed two semesters of eleven or more hours each during the academic year 1946-1947.

The results of the analysis are shown in Table 8. The percentage of veterans exceeding the average nonveteran was found to be 61; this advantage of the veterans is significant at the 2 per cent level.

7. *Turner University*

At least two special features characterize the analysis of data for freshmen at Turner University: first, the average grades are based on achievement-test scores used by the college rather than on the course grades; and second, students entering in the fall of 1945 have been combined with those entering in the fall of 1946 in making up the group. The inclusion

TABLE 9

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Turner University, College of Arts and Science

(Students who entered as freshmen in fall, [illegible] or fall, 1946)

I Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: Composite Test	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
Composite Test (converted score)	MV MN		.61 .60	66.3 68.1	7.3 8.0	100 101
First-Year College Average Grade	MV MN	.61 .60		2.47 2.73	.66 .81	100 101

Validity coefficient for combined group: .61

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	4.5403	1	Between .02 and .05
B. Equality of slopes	0.2406	1	Between .50 and .70
C. Equality of intercepts	3.7527	1	Between .05 and .10

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Nonveteran
Advantage expressed in grade units	0.16
Advantage expressed in standard error of estimate units	0.28
Per cent of veterans excelling the average nonveteran	39
Level of significance of difference (from IIC above)	Not significant

school course of study. As usual, freshmen in arts and science found that their program permitted a number of alternative ways of meeting various requirements; only English was a required course. For the larger group at Central State, the following definition may be stated: male students who entered as freshmen in the fall of 1946 and who completed two semesters of academic work during the academic year 1946-1947.

The following information may be useful in interpreting the results presented in Table 11. The Composite Test score was reported as a percentile rank based on six tests: Correctness and Effectiveness of Expression, Reading Materials in Social Studies, Reading Materials in Natural Science, Interpretation of Literary Materials, General Mathematical Ability, and General Vocabulary. The percentile scores were converted to a standard scale with a mean of 13 and a standard deviation of 4. The criterion was the First-Year College Average Grade. The grading system used five categories having the numerical values 4 to 0.

The results for students entering in the fall of 1946 are presented in Table 11. It will be noted that the Composite Test has a somewhat higher validity for veterans than for nonveterans, and that the validity of High-School Average is the same for both subgroups; in most institutions it has been found that high-school standing is more predictive of college grades for nonveterans than for veterans. Sixty-five per cent of the male veterans excel the average nonveteran, the most extreme difference so far encountered. The difference in freshman grades, when aptitude is considered, is highly significant.

The group of students which entered in 1945 may be defined as follows: male students who entered in the fall or winter semester of 1945, and who completed at least three semesters

TABLE 11

Comparison of Average Grades Earned by Veteran and Nonveteran Male Students

Central State University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: Composite Test	H. S. Avg. Grade	1st-Yr. Average Grade	Mean	*SD*	*N*
1. Composite Test (converted score)	MV		.48	.58	12.5	3.9	466
	MN		.43	.51	13.5	3.6	166
2. High-School Average Grade	MV	.48		.61	2.48	.62	466
	MN	.43		.61	2.74	.64	166
3. First-Year College Average Grade	MV	.58	.61		2.20	.68	466
	MN	.51	.61		2.19	.66	166

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .69; for MN, .67; and for the combined group, .68.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.016	1	Between .80 and .90
B. Equality of slopes	0.815	2	Between .50 and .70
C. Equality of intercepts	17.753	1	Less than .01

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.19
Advantage expressed in standard error of estimate units	0.39
Per cent of veterans excelling the average nonveteran	65
Level of significance of difference (from IIIC above)	1%

TABLE 12

Comparison of Average Grades Earned by Veteran and Nonveteran Male Students

Central State University, College of Arts and Science

(Students entering as freshmen during the academic year 1945–1946)

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with:		Mean	*SD*	*N*
		Composite Test	1st. 3 Sem. Average Grade			
Composite Test (converted score)	MV MN		.43 .49	13.5 14.5	3.8 4.2	135 59
First Three Semesters College Average Grade	MV MN	.43 .49		2.36 2.38	.53 .65	135 59

Validity coefficient for combined group: .45

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	2.792	1	Between .05 and .10
B. Equality of slopes	0.795	1	Between .30 and .50
C. Equality of intercepts	0.313	1	Between .50 and .70

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.04
Advantage expressed in standard error of estimate units	0.09
Per cent of veterans excelling the average nonveteran	54
Level of significance of difference (from IIC above)	Not significant

before the summer of 1947. An eight-week summer term was counted as one semester for three of the 135 veterans.

The analysis plan for the group entering in 1945 differed from that for the 1946 group in only two respects: First Three-Semesters College Average Grade was used as the criterion, and High-School Average was not included in the analysis of covariance. Basic results are shown in Table 12. The number of cases is rather small for the 1945 group; the results must therefore be interpreted with caution.

The validity coefficients for the Composite Test score were .43 and .49 for veterans and nonveterans respectively. Those for High-School Average turned out to be .33 and .65 for the two subgroups. Since the use of both predictors yielded regression planes which differed reliably in slope (at the 2 per cent level of significance), it was decided to use only the Composite Test as the predictor. From Table 12 it may be seen that 54 per cent of the veterans exceed the average nonveteran in average grade.

3. *Littletown State University*

Two groups, both entering as freshmen in the fall of 1946, were studied at Littletown State. Of these, the group entering the College of Arts and Science will be discussed in this section; the other group, which entered the College of Business, will be discussed in a later section. For the arts and science students, the analysis was limited to a comparison of the achievement of veteran and nonveteran students after allowing for differences in American Council Psychological Examination total scores.

Littletown State University is a coeducational state university located in a small midwestern city. Students are drawn mainly from the high schools of the state, and are admitted on the basis of high-school graduation. Stu-

TABLE 13

Comparison of Average Grades Earned by Veteran and Nonveteran Male Students

Littletown State University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
ACPE Total (raw score)	MV		.49	111	23	103
	MN		.34	115	20	107
First-Year College Average Grade	MV	.49		2.29	.68	103
	MN	.34		2.35	.69	107
Validity coefficient for combined group:			.42			

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.814	1	Between .30 and .50
B. Equality of slopes	0.464	1	Between .30 and .50
C. Equality of intercepts	0.00193	1	Between .95 and .98

III. Difference between Subgroups Ability Held Constant:

Superior subgroup	Male Nonveteran
Advantage expressed in grade units	0.00
Advantage expressed in standard error of estimate units	0.01
Per cent of veterans excelling the average nonveteran	50
Level of significance of difference (from IIC above)	Not significant

dents with poor averages in high school are admitted on a probationary basis. Freshmen in arts and science are required to take an English course; the remainder of their program is determined by broader requirements. The sample under consideration may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed eleven or more hours of work in the College of Arts and Science in each of the two semesters of the academic year, 1946-1947. The criterion, First-Year College Average Grade, was actually a point-hour ratio, based on a scale of five steps having the numerical values of 4 to 0. Raw scores on the ACPE were used.

The results are shown in Table 13. When allowance is made for ability differences the average grades of veterans and nonveterans at Littletown State are almost identical and, of course, the very slight difference in favor of nonveterans is not significant.

4. *Eastern City University*

For students entering Eastern City in the fall of 1946, the analysis was limited to a comparison of male veteran and nonveteran students, after allowing for differences in aptitude reflected in High-School Average Grade and Composite Score on entrance tests. The criterion measure was First-Year College Average Grade.

Eastern City is a coeducational municipal college of arts and science located in a large city. Students are admitted to Eastern City on a competitive basis, using primarily their high-school records and their scores on entrance examinations administered by the college. Although the freshman program in arts and science is not prescribed, a relatively large number of required courses, in English, mathematics, and social sciences, must be completed before graduation.

The groups selected for statistical analysis may be described as follows: male students who enrolled as beginning freshmen in the fall of 1946 and who completed two semesters of academic work. The limitation of the group to students entering in the fall greatly reduced the number of veterans available for study, since many veterans had been admitted during the

TABLE 14

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Eastern City University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: Composite Test+H. S. Average	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
Composite Test Score plus High-School Average	MV MN		.45 .54	165.5 167.5	4.6 6.5	53 147
First-Year College Average Grade	MV MN	.45 .54		0.26 0.37	.53 .59	53 147
Validity coefficient for combined group:			.53			

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.128	1	Between .70 and .80
B. Equality of slopes	0.0184	1	Between .80 and .90
C. Equality of intercepts	0.0414	1	Between .80 and .90

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Nonveteran
Advantage expressed in grade units	0.02
Advantage expressed in standard error of estimate units	0.04
Per cent of veterans excelling the average nonveteran	48
Level of significance of difference (from IIC above)	Not significant

previous spring. It was considered desirable, however, to exclude these veterans in order to obtain as comparable conditions as possible between veteran and nonveteran students.

For the statistical analysis, High-School Average and Composite Test score were combined, using equal weights. (The standard deviations were roughly equal.) The Composite Test score was based on performance on eight Cooperative tests: Mechanics of Expression, Effectiveness of Expression, Vocabulary, Speed of Comprehension, Level of Comprehension, Mathematics, Natural Science, and Social Studies. The results of the analysis are presented in Table 14. Forty-eight per cent of the veterans exceed the average nonveteran; the amount of overlapping of the distributions is very great, and, of course, the difference is not significant.

5. *Midwest City University*

Two groups, both entering as freshmen in 1946, were studied at Midwest City. The present section is concerned only with those entering the College of Arts and Science. For Midwest City students, the analysis was limited to a comparison of the relative achievement of male veteran and nonveteran students after eliminating the effect of differences in performance on the American Council Psychological Examination.

Midwest City is a large coeducational municipally supported institution. Students are admitted on the basis primarily of high-school record, being selected from students graduated in the upper two-thirds of their graduating classes. Freshmen in arts and science are required to take English and a survey course in western civilization and to fulfill additional broader requirements. The group studied may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed seven or more semester hours of academic work during each semester.

Total raw scores on the ACPE were used as the only predictor for this group. The criterion, the First-Year College Average Grade, is based on a scale of six points corresponding to the values 5 to 0. The results of the analysis are shown in Table 15.

An unusually large difference between the

TABLE 15

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Midwest City University, College of Arts and Science, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	83 *N*
ACPE Total (raw score)	MV		.57	120	23	
	MN		.29	120	23	83
First-Year College Average Grade	MV	.57		3.40	.77	72
	MN	.29		3.18	.80	72
Validity coefficient for combined group:			.44			

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	2.872	1	Between .05 and .10
B. Equality of slopes	3.323	1	Between .05 and .10
C. Equality of intercepts	3.622	1	Between .05 and .10

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.22
Advantage expressed in standard error of estimate units	0.31
Per cent of veterans excelling the average nonveteran	62
Level of significance of difference (from IIC above)	Not significant

validity coefficients for veterans and nonveterans was found for this group at Midwest City; the correlations are .57 and .29 for veterans and nonveterans respectively. Although 62 per cent of the veterans exceed the average nonveteran, the difference is not significant because of the small number of cases.

6. *Summary*

The results for the six groups from five state and municipal colleges of arts and science are in general similar to those found for the private institutions. For the private colleges, the veterans showed higher achievement in six out of seven cases, but the difference was significant (at the 2 per cent level) in only one case. For the public arts colleges, the veterans were superior in four out of six cases. The superiority of the veterans was significant at the 1 per cent level for one institution (Central State). At another (Western State) the results of the significance test were ambiguous because of difference in errors of estimate, which was significant at the 2 per cent level.

The tendency for high-school standing to have greater validity for nonveterans than for veterans is verified in one case, but at Central State the coefficient was the same for both subgroups.

C. Grades of Veteran and Nonveteran Students in Engineering Colleges of State and Municipal Universities

1. *Midwest Technological University*

At Midwest Technological University, a number of variables were studied in addition to those employed in the basic comparisons of veterans and nonveterans, which took into account High-School Average Grade and total score on the ACE Psychological Examination. For this group, attention was given to placement tests in English and mathematics, to the average grades and credits earned during each quarter, and to course grades in English and in mathematics.

Midwest Technological University is a land-grant college in a state which also supports a state university. Students who are residents of the state are admitted upon completion of an appropriate high-school program. Students in the different branches of engineering take a common freshman year, which includes chemistry, drawing, English, college algebra, trigonometry, and analytic geometry. The group selected for study may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed three quarters of engineering college work, with ten or more hours of academic work each quarter. The criterion, quarter average grade for the three semesters of the freshman college year, is based on a five-step scale with the numerical values 4 to 0.

Comparison of the means of veterans and nonveterans on the predictive measures showed that the average veteran excelled on only one predictor, the English test, although the differences were generally small. On the mathematics placement test, the advantage of the nonveterans amounted to about half of a standard deviation. Veterans were higher in college average grades during each quarter and in the courses studied separately, as well as in over-

TABLE 16

Comparison of Average Grades Earned by Veteran and Nonveteran Male Students

Midwest Technological University, College of Engineering, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with:			Mean	*SD*	*N*
		ACPE	H. S. Average Grade	1st-Yr. Average Grade			
1. ACPE (1945) Total (converted score)	MV MN		.44 .37	.60 .61	13.7 14.4	3.5 3.5	271 128
2. High-School Average Grade	MV MN	.44 .37		.62 .53	2.75 2.96	.62 .56	217 128
3. First-Year College Average Grade	MV MN	.60 .61	.62 .53		2.39 2.24	.63 .63	271 128

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .72; for MN, .76; and for the combined group, .70.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.601	1	Between .30 and .50
B. Equality of slopes	1.718	2	Between .30 and .50
C. Equality of intercepts	39.780	1	Less than .01

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.30
Advantage expressed in standard error of estimate units	0.70
Per cent of veterans excelling the average nonveteran	76
Level of significance of difference (from IIIC above)	1%

all average for the year. The mean number of credits was greater for nonveterans in each of the three quarters. Correlations between number of credits for each quarter and average grade for that quarter were found to be low and positive, for veterans and nonveterans separately, indicating a tendency for the better students to take heavier course loads.

The intercorrelations of the quarter averages ranged from .70 to .78 for nonveterans and from .68 to .75 for veterans, the average for each group being about .73. Assuming a reliability of .73 for the quarter averages, use of the Spearman-Brown formula would predict that the reliability of average grades based on the three quarters would be .89. When it is recalled that this method of estimating reliability tends to give an underestimate (because the three quarter grades are not random thirds of the freshman average grade), this figure may be considered satisfactorily high.

The results of the analysis of covariance are shown in Table 16. The multiple correlations of the ACPE in combination with High-School Average are unusually high: .72 and .76 for veterans and nonveterans respectively, and .70 for both subgroups combined. There is a marked tendency for veterans to earn higher grades than nonveterans of equal ability as measured by the predictors used. Seventy-six per cent of the veterans exceed the average nonveteran, which is the largest difference found in the study of twenty-five different groups.

2. *Middle State University*

At Middle State, a number of predictors and criteria were studied. Here, American Council Psychological Examination total score, High-School Rank, Year of High-School Graduation, First-Year College Average Grade, and course grades in drawing and mathematics were ana-

TABLE 17

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Middle State University, College of Engineering, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	H. S. Rank	1st-Yr. Average Grade	Mean	*SD*	*N*
1. ACPE (1937) Total (raw score)	MV		.46	.42	88.1	21.5	352
	MN		.49	.28	93.1	22.3	98
2. High-School Rank (converted score)	MV	.46		.51	15.6	3.1	352
	MN	.49		.53	16.4	3.3	98
3. First-Year College Average Grade (4-step scale)	MV	.42	.51		1.40	.66	352
	MN	.28	.53		1.32	.60	98

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .56; for MN, .53; and for the combined group, .53.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.913	1	Between .30 and .50
B. Equality of slopes	4.702	2	Between .05 and .10
C. Equality of intercepts	8.816	1	Less than .01

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.19
Advantage expressed in standard error of estimate units	0.34
Per cent of veterans excelling the average nonveteran	63
Level of significance of difference (from IIIC above)	1%

TABLE 18

COMPARISON OF DRAWING GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Middle State University, College of Engineering, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with:			Mean	*SD*	*N*
		ACPE	H. S. Rank	Drawing Grade			
1. ACPE (1937) Total (raw score)	MV MN		.46 .49	.23 .18	88.1 93.1	21.5 22.3	352 98
2. High-School Rank (converted score)	MV MN	.46 .49		.27 .14	15.6 16.4	3.1 3.3	352 98
3. Drawing Grade—First Quarter (5-step scale)	MV MN	.23 .18	.27 .14		0.71 0.53	.83 .91	352 98

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .30; for MN, .19; and for the combined group, .26.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	2.197	1	Between .10 and .20
B. Equality of slopes	1.492	2	Between .30 and .50
C. Equality of intercepts	6.849	1	Less than .01

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.25
Advantage expressed in standard error of estimate units	0.30
Per cent of veterans excelling the average nonveteran	62
Level of significance of difference (from IIIC above)	1%

lyzed. A unique feature of the analysis of Middle State data is the comparison of veterans with nonveterans not only on the basis of First-Year Average Grade, but also on their grades in mathematics and drawing. In all three comparisons, differences in High-School Rank and in ACPE total score were taken into account.

Middle State is a large state university located in a midwestern city. Students of engineering are drawn from graduates ranking in the upper half of their high-school classes. Chemistry, English, mathematics, and drawing are required courses for engineering students. The group studied may be defined as follows: students who were admitted to the College of Engineering as freshmen in the fall of 1946 and who completed three quarters of work, including eleven or more hours each quarter.

The predictors employed in the study included total score on the 1937 edition of the American Council Psychological Examination (expressed in terms of raw score) and High-School Rank. The ACPE was given to high-school students in their senior year throughout the state. Rank in high-school class was converted from percentile ranks to a standard scale with a mean of 13 and a standard deviation of 4.

The principal criterion employed is the First-Year College Average Grade based on a four-category scale having the equivalent numerical values of 3 to 0. First-quarter grades in engineering, drawing, and mathematics were used as additional criteria. For these variables, a five-step grading scale was used, the values ranging from 3 to —1. Still another variable, Year of High-School Graduation, was used.

Study of the year of high-school graduation showed that most nonveterans were graduated in 1946, while the typical veteran last attended high school in 1943. Veterans, of course, showed much greater variability with respect to this measure. For veterans the correlations involving year of graduation range from .20 (with High-School Rank) to —.01 (with Drawing

Grade). Except for the r of —.01, the remaining four correlations are positive and above .08, which indicates a very slight tendency for the more recent high-school graduates to be superior. Such a relationship could easily result from admissions policy or a process of self-selection and does not necessarily reflect a general tendency.

The best predictor of First-Year Average Grade is High-School Rank. The ACPE yields a validity coefficient for veterans which is considerably higher than for nonveterans; the coefficients are .42 and .28. As might be expected, the Drawing Grades are not well predicted by either of the predictive measures used. Validity coefficients for Mathematics Grades are somewhat higher and are of about the same magnitude for veterans and nonveterans.

It may be noted that the correlations between Mathematics Grades and First-Year College Average Grades were .75 for nonveterans and .76 for veterans; the corresponding figures for Drawing Grades with Average Grades were .35 and .50. The correlations between Drawing Grades and Mathematics Grades were .19 for nonveterans and .30 for veterans.

Three analyses of covariance were computed for the data at Middle State University, one for each of the criteria employed. The results for First-Year College Average Grade are shown in Table 17. The advantage of veteran students in First-Year Average Grade at the College of Engineering at Middle State University is highly significant.

Table 18 presents the results where Drawing Grade is used as the criterion. This criterion is less predictable than either of the others; the three multiple correlations are .30, .19, and .26. The intercepts are significantly different at the 1 per cent level.

The results of the analysis of covariance in which Mathematics Grade is used as the cri-

TABLE 19

COMPARISON OF MATHEMATICS GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Middle State University, College of Engineering, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	H. S. Rank	Math. Grade	Mean	*SD*	*N*
1. ACPE (1937) Total (raw score)	MV		.46	.34	88.1	21.5	352
	MN		.49	.28	93.1	22.3	98
2. High-School Rank (converted score)	MV	.46		.41	15.6	3.1	352
	MN	.49		.40	16.4	3.3	98
3. Mathematics Grade—First Quarter (5-step scale)]	MV	.34	.41		0.49	1.10	352
	MN	.28	.40		0.39	1.10	98

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .44; for MN, .41; and for the combined group, .43.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.0566	1	Between .80 and .90
B. Equality of slopes	0.725	2	Between .50 and .70
C. Equality of intercepts	4.342	1	Between .02 and .05

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.24
Advantage expressed in standard error of estimate units	0.24
Per cent of veterans excelling the average nonveteran	59
Level of significance of difference (from IIIC above)	5%

TABLE 20

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Midwest City University, College of Engineering, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Subgroup	Correlation with: ACPE	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
ACPE Total (raw score)	MV MN		.45 .36	118 115	19 20	167 171
First-Year College Average Grade	MV MN	.45 .36		3.34 3.32	.74 .77	167 171

Validity coefficient for combined group: .40

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	1.184	1	Between .20 and .30
B. Equality of slopes	0.646	1	Between .30 and .50
C. Equality of intercepts	0.101	1	Between .70 and .80

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Nonveteran
Advantage expressed in grade units	0.02
Advantage expressed in standard error of estimate units	0.03
Per cent of veterans excelling the average nonveteran	49
Level of significance of difference (from IIC above)	Not significant

terion are shown in Table 19. The intercepts are significantly different at the 5 per cent level in the case of this criterion.

For all three criteria used at Middle State, then, the veterans proved to achieve significantly higher grades than nonveterans when allowance is made for the difference in ability. When Mathematics Grade is used as the criterion, however, the difference is significant at only the 5 per cent level.

3. *Midwest City University*

The analysis of data for engineering students at Midwest City was limited to a comparison of their relative first-year grades after allowing for differences in total score on the American Council Psychological Examination.

Like the students in arts and science, Midwest City engineering students are drawn from the upper two-thirds of their high-school graduating classes. As freshmen, they take English, drawing, descriptive geometry, chemistry, college algebra, trigonometry, analytic geometry, and mechanics. The engineering program at Midwest City is well known for its practical emphasis. The sample upon which the analysis was based may be defined as follows: students entering as beginning freshmen in the fall term of 1946 who completed the full academic year 1946-1947.

The results of the analysis are shown in Table 20. In the College of Engineering at Midwest City the veteran and nonveteran students are unusually similar with respect to college achievement in relation to aptitude.

4. *Southern Technological University*

Analysis of data for engineering students at Southern Tech was limited to a comparison of the relative achievement of male veteran and nonveteran students after allowing for differences in total score on the American Council Psychological Examination. In order to secure enough cases to warrant analysis, it was necessary to combine students who entered in the fall of 1945 with those who entered in the fall of 1946.

Southern Tech is the land-grant college for

TABLE 21

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Southern Technological University, College of Engineering

(Students who entered as freshmen in fall, 1945, or fall, 1946)

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
ACPE Total (raw score)	MV		.48	111	21	120
	MN		.41	109	21	50
First-Year College Average Grade	MV	.48		2.50	.70	120
	MN	.41		2.22	.60	50

Validity coefficient for combined group: .46

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.956	1	Between .30 and .50
B. Equality of slopes	0.885	1	Between .30 and .50
C. Equality of intercepts	6.110	1	Between .01 and .02

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.25
Advantage expressed in standard error of estimate units	0.42
Per cent of veterans excelling the average nonveteran	66
Level of significance of difference (from IIC above)	2%

white students in a state which also supports a state university. Students are admitted upon completion of an appropriate high-school course. Engineering students carry a freshman program which includes chemistry, English, algebra, trigonometry, analytic geometry, drawing, descriptive geometry, and shop. The group studied may be defined as follows: students who entered as beginning freshmen either in the fall of 1945 or the fall of 1946 and who completed three quarters of engineering work (including eleven or more credits per quarter) during the academic year in which they entered. Students who received ten or more credits in specific subjects for training received while in the armed services were not considered to be "beginning" freshmen and were excluded.

The following information may aid in interpreting the results presented in Table 21: the criterion, First-Year College Average Grade, is based on the work of the three quarters of the freshman year. Grades are based on a five-step scale of 4 to 0. The only predictor used was the total raw score on the American Council Psychological Examination. The hypotheses of equal errors of estimate and equal slopes are not disproved, but the difference in intercepts is significant at the 2 per cent level.

5. *Summary*

In three of the four engineering schools studied, veterans were found to earn higher freshman average grades than nonveterans, when the influence of ability differences is eliminated. The difference found at two of the universities would be expected to occur by chance less than once in a hundred times, and at the third institution less than twice in a hundred. At the remaining college there was practically no difference between veteran and nonveteran students, although what difference occurred was in favor of the nonveterans. At one university, the veterans were also found to earn better grades than nonveterans of the same ability level in two courses, drawing and mathematics.

The tendency was found at Midwest Tech for the abler students, as judged both by predictor variables and college grades, to take the heavier course loads. Nonveterans tended to take slightly heavier course loads than the veterans.

High-school standing has usually been found to predict grades better for nonveteran than veteran students, but at Midwest Tech the opposite conclusion was reached, and at Middle State the difference in validity was small.

D. Grades of Veteran and Nonveteran Students in Agriculture and Business Colleges in State Universities

1. *Midwest Technological University*

For students in the College of Agriculture at Midwest Tech, a fairly extensive analysis was undertaken. Intercorrelations of the following variables were studied: part and total scores on the American Council Psychological Examination, score on an English placement test, High-School Average, academic load carried for each quarter of the freshman college year, average grade for each quarter, and course grades in English. For the comparison of relative achievement between male veterans and male nonveterans, allowance was made for differences in total score on the ACPE and in High-School Average Grade; the criterion was First-Year College Average Grade.

Residents of the state may be admitted to the College of Agriculture at Midwest Tech upon the completion of an appropriate high-school course. English is the only course required for all agriculture students, although various other courses are required for certain large subgroups of these students. Even during the freshman year, a definite emphasis upon agricultural subjects is present, and students preparing for different specialties take somewhat different programs. The group upon which the analysis was based may be defined as follows: students who entered as beginning freshmen in the fall of 1946 and who completed three quarters of work in the College of Agriculture (including ten or more hours of work each term) during the academic year 1946-1947.

It was found that veterans tended to take slightly lighter course loads than the nonveterans, the difference amounting to about half a credit-hour. Number of credits correlated positively with all the other measures, both predictors and criteria, but no systematic difference is apparent between veterans and nonveterans in the magnitude of the correlations. As in the other analyses where information on this point was obtained, the better students tend to take the heavier course loads. Comparing the means of the predictors, we find that the nonveterans are higher only on the ACPE Quantitative score and on High-School Average. Without exception the veterans exceed the nonveterans on the criterion variables.

Somewhat different results are obtained for veterans and nonveterans with regard to the validity of the predictors. For nonveterans, High-School Average is unquestionably the best predictor of quarter average grades, while for veterans the English test appears to be generally superior. For predicting grades in English, the English test is again relatively better for veterans than for nonveterans.

The analysis of covariance results are shown in Table 32. The ACPE and High-School Average were selected as the predictors, and the criterion was First-Year College Average Grade. The standard errors of estimate were significantly different at the 2 per cent level; grades of nonveterans tended to be more predictable. The equality of the slopes of the regression planes was not disproved, but the intercepts were found to be significantly different at the 1 per cent level. The test of the significance of the difference in intercepts is difficult to interpret in the light of the outcome of the test of Hypothesis A.

2. *Midwest State University*

Analysis of academic data for students in the College of Business of Midwest State University was limited to a comparison of achievement after allowing for differences in scores on the Ohio State University Psychological Examination.

Midwest State is a large coeducational state university located in a large city. Students who are residents of the state are admitted upon the completion of an appropriate high-school course of study. Students of business are required to take courses in English, economics, business, and geography, and to meet certain broader requirements during their freshman year. The group to be studied is defined as follows: students who entered as beginning freshmen

TABLE 22

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Midwest Technological University, College of Agriculture, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with:			Mean	*SD*	*N*
		ACPE	H. S. Average Grade	1st-Yr. Average Grade			
1. ACPE (1945) Total (converted score)	MV		.38	.52	11.9	3.9	140
	MN		.49	.48	11.5	3.5	102
2. High-School Average Grade	MV	.38		.56	2.52	.64	140
	MN	.49		.68	2.75	.59	102
3. First-Year College Average Grade	MV	.52	.56		2.30	.60	140
	MN	.48	.68		2.15	.51	102

II. Multiple Correlations (Variables 1 and 2 vs. Variable 3): for MV, .65; for MN, .70; and for the combined group, .64.

III. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	5.425	1	Between .01 and .02
B. Equality of slopes	2.357	2	Between .30 and .50
C. Equality of intercepts	16.137	1	Less than .01

IV. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.23
Advantage expressed in standard error of estimate units	0.54
Per cent of veterans excelling the average nonveteran	71
Level of significance of difference (from IIIC above)	Ambiguous

in the fall of 1946 and who completed three quarters in the College of Business, including at least ten hours of work in each quarter. Students who were granted ten or more hours of credit in specific courses for training received during service were excluded from the group. The results for Midwest State business students are presented in Table 23.

The Ohio State University Psychological Examination (OSPE) proved to be a good predictor of freshman grades for both veteran and nonveteran students; the validity coefficients were .59 and .55. (The OSPE scores were converted to a standard scale having a mean of 13 and a standard deviation of 4.) Since high-school standing was reported in very coarse groupings (upper, middle, and lower third of the high-school class), it was not thought worth while to include it as a predictor. The criterion, which as usual was the First-Year College Average Grade, was based on a scale of five categories having the values of 4 to 0. The errors of estimate and slopes of the regression lines did not prove to be significantly different, but the intercepts were significantly different at the 5 per cent level.

3. *Littletown State University*

For students in the College of Business at Littletown State, analysis was limited to a comparison of the achievement of male veteran and nonveteran students, taking into account differences in total scores on the American Council Psychological Examination. The criterion again is First-Year College Average Grade.

Residents of the state were admitted to the College of Business upon completion of an appropriate course of study in high school. Those with poor high-school records were admitted on somewhat of a probationary basis.

TABLE 23

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Midwest State University, College of Business, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: OSPE	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
OSPE (converted score)	MV		.59	12.0	3.2	232
	MN		.55	13.2	4.0	58
First-Year College Average Grade	MV	.59		2.26	.65	232
	MN	.55		2.22	.70	58

Validity coefficient for combined group: .57

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	1.201	1	Between .20 and .30
B. Equality of slope	1.189	1	Between .20 and .30
C. Equality of intercepts	5.029	1	Between .02 and .05

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.18
Advantage expressed in standard error of estimate units	0.33
Per cent of veterans excelling the average nonveteran	63
Level of significance of difference (from IIC above)	5%

Students of business were required to take English and a survey course in business and to meet certain broader requirements during their freshman year. The group selected for study may be defined as follows: male students who entered as beginning freshmen in the fall of 1946 and who completed two semesters, including eleven or more hours of credit in each.

The results are presented in Table 24. No significant differences were found, in the analysis of covariance, although 59 per cent of the veterans exceeded the average nonveteran student in the College of Business at Littletown State.

4. *Summary*

Male veterans proved to be superior to male nonveterans with respect to college achievement (in relation to ability) at the single agriculture college and in both business colleges studied. The difference was significant at the 5 per cent level in one instance (Midwest State College of Business). The difference was not significant in the case of the business students at Littletown State, and was ambiguous for agriculture students at Midwest Tech.

The tendency for veterans to take lighter course loads than nonveterans was again found for agriculture students at Midwest Tech, and again the tendency was noted for the more able students to take heavier loads within each subgroup. High-school standing was again found to yield better prediction of grades for male nonveterans than for veterans.

TABLE 24

COMPARISON OF AVERAGE GRADES EARNED BY VETERAN AND NONVETERAN MALE STUDENTS

Littletown State University, College of Business, Freshmen, 1946–1947

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: ACPE	Correlation with: 1st-Yr. Average Grade	Mean	*SD*	*N*
ACPE Total (raw score)	MV		.55	106	23	142
	MN		.57	107	21	65
First-Year College Average Grade	MV	.55		2.34	.65	142
	MN	.57		2.23	.61	65
Validity coefficient for combined group:			.56			

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.347	1	Between .50 and .70
B. Equality of slopes	0.0191	1	Between .80 and .90
C. Equality of intercepts	2.436	1	Between .10 and .20

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.12
Advantage expressed in standard error of estimate units	0.23
Per cent of veterans excelling the average nonveteran	59
Level of significance of difference (from IIC above)	Not significant

E. GRADES OF INTERRUPTED VETERANS AND UNINTERRUPTED NONVETERANS

In the studies so far reported, the first-year average grade has been used as the primary criterion, and ability to do college work has been assessed on the basis of one or more tests of scholastic aptitude, tool skills, or achievement in high-school subjects, used wherever possible in combination with a measure of high-school standing. In the studies of interrupted veterans, however, the veteran is in a sense compared with himself as a nonveteran. The interrupted veteran, as defined here, is a student who attended college for a certain period of time (usually one year) as an ordinary civilian student, after which he entered military service; upon completion of his military service, he re-entered the same institution and completed a defined amount of academic work (usually one year). For such students it seemed that the best basis for evaluating the effect of the student's intervening experiences on college achievement was a comparison of his achievement as a civilian student (before war service) with his achievement as a veteran student.

In planning the analysis, it was thought likely that grades earned during the term just before the student entered the armed forces might not be typical, since the prospects of leaving college shortly might lead the student to slacken his efforts in academic work. It was also supposed that grades for the first term following the return to college might not be typical, partly because background information and study skills might need refreshing before a student regained his original effectiveness, and partly because students returning to college might be temporarily fired with an enthusiasm which would not persist. In three of the five groups studied, the first-term grade was used as the predictor and the fourth-term grade as the criterion, leaving out of consideration the doubtful second and third terms. The analysis of covariance method was again used, in order to compare the second-term sophomore grades of veterans and nonveterans, taking account of ability as indicated by grades earned in the first term of the freshman year.

1. *Adams University*

A relatively elaborate study was made of the group containing interrupted veterans at Adams University, in order to obtain a more complete picture of the relationships of a number of pertinent variables. In this study, the typical definition of groups was made: veterans completed one year before war service and one year afterward; nonveterans entered in the fall of 1945 and completed two consecutive years of college work. Students not enrolled in the arts and science curriculum, and veterans given advanced standing for college training during their service were excluded.

The intercorrelations obtained for the two

TABLE 25

INTERCORRELATIONS OF SCHOLASTIC APTITUDE TEST SCORES, ADJUSTED SCHOOL RANK, PREDICTED GRADE, AND FIRST- AND SECOND-YEAR COLLEGE AVERAGE GRADES

Adams University, College of Arts and Science

(Veterans who completed one year before war service and one year after war service; nonveterans who entered in the fall of 1945 and completed two years.)

		Male Nonveterans (N=111)									
		SAT-V	SAT-M	Adjusted School Rank	Predicted Grade	1st-Sem. Average	2nd-Sem. Average	3rd-Sem. Average	4th-Sem. Average	Mean	S. D.
Interrupted Veterans (N=134)	SAT-Verbal		.33	.26	.72	.60	.53	.47	.46	559	107
	SAT-Mathematics	.28		.24	.55	.42	.23	.32	.22	574	93
	Adjusted School Rank	.31	.29		.62	.49	.41	.46	.48	74.7	8.2
	Predicted Grade	.67	.46	.73		.70	.63	.64	.64	74.5	7.3
	1st-Semester Average	.60	.34	.43	.63		.80	.72	.66	75.6	7.8
	2nd-Semester Average	.54	.22	.49	.66	.78		.73	.71	76.4	7.4
	3rd-Semester Average	.46	.39	.37	.55	.55	.52		.83	76.9	6.8
	4th-Semester Average	.48	.22	.31	.51	.59	.57	.80		78.0	6.7
	Mean	547	576	74.7	72.8	72.9	72.8	77.6	78.6		
	Standard Deviation	101	78	7.5	6.6	8.1	7.8	6.0	6.3		

subgroups are shown in Table 25. The statistics above and to the right of the diagonal were obtained for the nonveteran subgroup, and those below the diagonal for the interrupted veterans.

The hypothesis that First-Semester College Average Grades would furnish a better prediction of Fourth-Semester Grades than tests plus high-school standing was verified. For veterans, the Predicted Grade, which was the best of the ordinary predictors, correlated .51 with Fourth-Semester Average; the correlation of First-Semester Average with Fourth-Semester Average was .59. For the nonveteran subgroups, however, the difference was much less marked; the analogous correlations were .64 and .66.

It is of interest to note that the validity coefficients of the two scholastic aptitude tests against the fourth-term averages are rather similar for veterans and nonveterans, but the validities for Adjusted School Rank and Predicted Grade are noticeably lower after the interruption of training than during the initial college year. This finding is consistent with the previous observation that high-school standing ordinarily has less predictive value for veterans than for nonveterans.

It is also worthy of note that the intercorrelations of term grades which cross the time of interruption (i.e., all except first-semester vs. second-semester and third-semester vs. fourth-semester) are considerably lower for veterans than for nonveterans. This suggests that military service or other related experiences have affected the academic performance of the veteran students.

The means for the veterans are slightly lower on most of the predictive measures, and their First- and Second-Semester Average Grades are lower than for nonveterans. The average grades for the third and fourth semesters, on the other hand, are higher for veterans. It will be

TABLE 26

COMPARISON OF AVERAGE FOURTH-SEMESTER GRADES EARNED BY INTERRUPTED VETERAN AND UNINTERRUPTED NONVETERAN MALE STUDENTS

Adams University, College of Arts and Science

(Veterans who completed one year before war service, one year after war service; nonveterans who entered in the fall of 1945 and completed two years.)

I. Correlations, Means, and Standard Deviations:

Variable	Subgroup	Correlation with: First-Sem. Average Grade	Correlation with: Fourth-Sem. Average Grade	Mean	*SD*	*N*
First-Semester College Average Grade	MV MN		.59 .66	72.9 75.6	8.1 7.8	134 111
Fourth-Semester College Average Grade	MV MN	.59 .66		78.6 78.0	6.3 6.7	134 111

Validity coefficient for combined group: .61

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.00474	1	Between .90 and .95
B. Equality of slopes	1.822	1	Between .10 and .20
C. Equality of intercepts	8.298	1	Less than .01

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	1.92
Advantage expressed in standard error of estimate units	0.38
Per cent of veterans excelling the average nonveteran	65
Level of significance of difference (from IIC above)	1%

noted that the mean Predicted Grade for veterans was 72.8 and the mean First-Semester Average was almost the same: 72.9. The nonveterans exceeded the prediction slightly more; the predicted mean grade was 74.5 and the mean of the First-Semester Average Grade was 75.6. In the second term, the veterans' mean was practically unchanged, while the nonveterans' mean grade increased slightly. There is then some slight evidence for believing that the anticipation of being drafted did lower the relative achievement of veterans to some extent.

The analysis of covariance results, comparing the regressions of Fourth-Semester Grades on First-Semester Grades for interrupted veterans and ordinary nonveterans, are shown in Table 26. The hypotheses of equal errors of estimate and equal slopes are not disproved, but the intercepts of the regression lines are found to be significantly different at the 1 per cent level. Sixty-five per cent of the interrupted veterans excel the average nonveteran in Fourth-Semester Grade when adjustment is made for ability differences as measured by First-Semester Average Grade.

This result is particularly striking in view of the results previously reported for the freshman veteran and nonveteran students at Adams; for the freshmen, no significant difference was found, although a relatively small difference would have been significant because so many students were included in the analysis. In the freshman group, only 52 per cent of the veterans excelled the average nonveteran.

2. *Stewart University*

At Stewart, the only variables taken into consideration were the average grades earned during the first four semesters of college work. As at Adams, the study involved a comparison of nonveteran students with veterans who had completed two full semesters of college work in the arts and science college, who were interrupted for military service, and who subsequently completed the two semesters of the sophomore year following their return from military duty. The nonveterans chosen for comparison entered in the fall of 1945, and completed two full years of college work. Transfer students and students given advanced standing for military training were eliminated. The predictor used was the First-Semester Average Grade, and the criterion was Fourth-Semester Average Grade. The grading system involved seven categories which were given the values of 7 to 1 for purposes of analysis.

The intercorrelations, means, and standard deviations of the four semester average grades for nonveterans and interrupted veterans are shown in Table 27. The intercorrelations are uniformly lower for veterans than for nonveterans, but the differences are especially large for

TABLE 27

INTERCORRELATIONS OF FIRST- AND SECOND-YEAR COLLEGE AVERAGE GRADES

Stewart University, College of Arts and Science

(Veterans who completed one year before war service and one year after war service; nonveterans who entered in the fall of 1945 and completed two years.)

		Male Nonveterans ($\underline{N}$=70)					
		1st-Sem. Average	2nd-Sem. Average	3rd-Sem. Average	4th-Sem. Average	Mean	S. D.
Interrupted Veterans ($\underline{N}$=55)	First-Semester College Average		.79	.73	.68	5.04	.99
	Second-Semester College Average	.74		.77	.72	4.91	.85
	Third-Semester College Average	.50	.46		.77	4.94	.92
	Fourth-Semester College Average	.67	.64	.70		5.13	.82
	Mean	4.88	4.77	5.29	5.38		
	Standard Deviation	.86	.95	.71	.68		

TABLE 28

COMPARISON OF AVERAGE FOURTH-SEMESTER GRADES EARNED BY INTERRUPTED VETERAN AND UNINTERRUPTED NONVETERAN MALE STUDENTS

Stewart University, College of Arts and Science

(Veterans who completed one year before war service, one year after war service; nonveterans who entered in the fall of 1945 and completed two years.)

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with:		Mean	*SD*	*N*
		First-Sem. Average Grade	Fourth-Sem. Average Grade			
First-Semester College Average Grade	MV MN		.67 .68	4.88 5.04	.86 .99	55 70
Fourth-Semester College Average Grade	MV MN	.67 .68		5.38 5.13	.68 .82	55 70

Validity coefficient for combined group: .65

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	1.756	1	Between .10 and .20
B. Equality of slopes	0.0489	1	Between .80 and .90
C. Equality of intercepts	11.104	1	Less than .01

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.32
Advantage expressed in standard error of estimate units	0.57
Per cent of veterans excelling the average nonveteran	72
Level of significance of difference (from IIC above)	1%

First-Semester vs. Third-Semester and Second-Semester vs. Third-Semester Average; the correlations are much lower for interrupted veterans than for the nonveterans, presumably because of influences related to the period of war service. It is interesting to note, however, that the correlations of First- and Second-Semester Averages with Fourth-Semester Average, which also bridge the period of interruption for veterans, do not show nearly as great a difference.

The means of veterans' grades are lower for both first and second semesters, but higher for third and fourth semesters than the grades of nonveterans. Both subgroups show a decrease in mean grade from the first semester to the second, and the difference is nearly the same for both subgroups. Apparently, either the anticipation of induction did not result in a letdown for the veterans, or the letdown was equally influential during both terms of the freshman year.

The magnitude of the intercorrelations between adjacent semester averages suggests that the grades are reasonably reliable.

Turning to Table 28, it will be seen that the correlations between the two semester grades employed as predictor and criterion were high: .67 for veterans and .68 for nonveterans. The analysis of covariance shows that differences in errors of estimate and regression slopes are not significant, while the difference in intercepts of the regression lines is significant at the 1 per cent level. Seventy-two per cent of the veterans exceed the average nonveteran in Fourth-Semester Grades when appropriate allowance is made for differences in First-Semester Grades.

3. *Eastern City University*

At Eastern City, the analysis was concerned only with First- and Fourth-Semester College Average Grades in arts and science. The groups studied were defined in the same manner as at Adams and Stewart. Uninterrupted nonveterans

TABLE 29

COMPARISON OF AVERAGE FOURTH-SEMESTER GRADES EARNED BY INTERRUPTED VETERAN AND UNINTERRUPTED NONVETERAN MALE STUDENTS

Eastern City University, College of Arts and Science

(Veterans who completed one year before war service, one year after war service; nonveterans who entered in the fall of 1945 and completed two years.)

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: First-Sem. Average Grade	Correlation with: Fourth-Sem. Average Grade	Mean	*SD*	*N*
First-Semester College Average Grade	MV		.49	0.19	.68	70
	MN		.60	0.29	.65	99
Fourth-Semester College Average Grade	MV	.49		0.57	.58	70
	MN	.60		0.33	.60	99

Validity coefficient for combined group: .53

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.357	1	Between .50 and .70
B. Equality of slopes	1.374	1	Between .20 and .30
C. Equality of intercepts	13.779	1	Less than .01

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.29
Advantage expressed in standard error of estimate units	0.59
Per cent of veterans excelling the average nonveteran	72
Level of significance of difference (from IIC above)	1%

who entered in 1945 and who completed the sophomore year comprised the nonveteran subgroup; the interrupted veterans completed the freshman year before military service and the sophomore year after military service.

For the analysis of covariance, First-Semester Average Grade was used as the predictor; Fourth-Semester Average Grade served as the criterion. The grading system involved five categories having values +2 to −2. Transfer students and students enrolled for fewer than twelve credits in any of the four semesters were eliminated.

The relationship between First- and Fourth-Semester Grades was not as marked at Eastern City as at Adams and Stewart; the correlations were .49 and .60 for the two subgroups, as shown in Table 29. A similar pattern of mean grades was found, however. The results of the analysis of covariance are also similar to that of the other two institutions. The hypotheses of equal standard errors and equal slopes are not disproved, while the difference in intercepts is found to be significant at the 1 per cent level. Seventy-two per cent of the interrupted veterans are found to exceed the average nonveteran student.

4. *Midwest Technological University, College of Engineering*

At Midwest Tech, the nonveteran comparison group was composed of prewar students rather than a postwar group as in the three studies just discussed. In this study, it was possible to increase the size of the interrupted group considerably by introducing some flexibility into the defined pattern of interruption. In the three studies just discussed, the interruption occurred between the second and third semesters. At Midwest Tech a veteran whose college career was interrupted at any time after the second quarter and before the seventh quarter was considered to fall within the defined pattern, provided that he completed at least eight quarters. As in the other studies of interrupted veterans, students who were given advanced standing for college training received as members of the armed forces were excluded. The nonveterans entered in the fall of 1939 and completed eight quarters of work; they were, of course, not interrupted. Transfer students and students who changed divisions within the university during the eight quarters were excluded from the study. Students who enrolled for fewer than 10 hours of work in any quarter or fewer than 12 hours in the first, second, or eighth quarter were also excluded.

A summer term counted as a regular quarter, provided the student was registered for twelve or more credit hours. Students were also rejected if they had not taken freshman mathematics or if they lacked a record of high-school standing. The same procedures described here apply to Midwest Tech College of Agriculture students; the study of the agriculture students will be described in the following section.

In the analysis of the data, First- and Second-Quarter College Average Grades were pooled and used as a single predictor; Eighth-Quarter Average Grades were used as the criterion. The results of the analysis are shown in Table 30. The validity coefficients tend to be somewhat lower than were found for the studies involving interrupted veterans at Adams and Stewart, probably because the criterion grades were earned at a time which is more remote from the predictor grades. The veterans were again found to earn lower grades during the initial two quarters (the predictor) than the nonveterans; on the Eighth-Quarter Grades (the criterion) the relative position of the two groups is reversed. The errors of estimate and slopes of the regression lines are not significantly dif-

TABLE 30

COMPARISON OF AVERAGE EIGHTH-QUARTER GRADES EARNED BY INTERRUPTED VETERAN AND UNINTERRUPTED NONVETERAN STUDENTS

Midwest Technological University, College of Engineering

(Veterans who completed two to six quarters before war service and two to six quarters after war service for a total of at least eight quarters; nonveterans who entered in the fall of 1939 and completed eight quarters.)

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: 1 & 2 Quarter Average Grade	Correlation with: 8th-Quarter Average Grade	Mean	*SD*	*N*
First- and Second-Quarter College Average Grade	MV		.57	2.26	.72	140
	MN		.45	2.37	.64	215
Eighth-Quarter College Average Grade	MV	.57		2.43	.74	140
	MN	.45		2.38	.67	215
Validity coefficient for combined group:			.50			

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	0.0731	1	Between .70 and .80
B. Equality of slopes	1.496	1	Between .20 and .30
C. Equality of intercepts	2.640	1	Between .10 and .20

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.11
Advantage expressed in standard error of estimate units	0.18
Per cent of veterans excelling the average nonveteran	57
Level of significance of difference (from IIC above)	Not significant

TABLE 31

COMPARISON OF AVERAGE EIGHTH-QUARTER GRADES EARNED BY INTERRUPTED VETERAN AND UNINTERRUPTED NONVETERAN MALE STUDENTS

Midwest Technological University, College of Agriculture

(Veterans who completed two to six quarters before war service and two to six quarters after war service for a total of at least eight quarters; nonveterans who entered in the fall of 1939 and completed eight quarters.)

I. Correlations, Means, and Standard Deviations:

Variable	Sub-group	Correlation with: 1 & 2 Quarter Average Grade	Correlation with: 8th-Quarter Average Grade	Mean	*SD*	*N*
First- and Second-Quarter College Average Grade	MV MN		.57 .62	2.16 2.30	.63 .60	57 106
Eighth-Quarter College Average Grade	MV MN	.57 .62		2.62 2.44	.69 .59	57 106

Validity coefficient for combined group: .57

II. Analysis of Covariance Results:

Hypothesis	Chi-square	Degrees of Freedom	Probability that a value greater than obtained would arise by chance
A. Equality of errors of estimate	2.901	[illegible]	Between .05 and .10
B. Equality of slopes	0.0000	1	Greater than .99
C. Equality of intercepts	9.255	1	Less than .01

III. Difference between Subgroups with Ability Held Constant:

Superior subgroup	Male Veteran
Advantage expressed in grade units	0.26
Advantage expressed in standard error of estimate units	0.51
Per cent of veterans excelling the average nonveteran	69
Level of significance of difference (from IIC above)	1%

ferent. The difference in intercepts is also found to be within the range of chance expectancy; 57 per cent of the interrupted veterans excel the average nonveteran. This engineering group at Midwest Tech is the only group among the five involving interrupted veterans where college achievement in relation to ability did not yield a statistically significant difference in intercepts.

5. *Midwest Technological University, College of Agriculture*

For Midwest Tech students of agriculture, the definition of the veteran and nonveteran groups and the plan of the analysis were in all respects parallel to those for the students of engineering described in the previous section.

Results of the analysis for this group are shown in Table 31. The differences in means show the same trend as did the preceding groups involving interrupted veterans: the veterans earned lower grades on the predictor and higher grades on the criterion measure. The advantage of the interrupted veterans is found to be significant at the 1 per cent level. Sixty-nine per cent of the interrupted veterans excel the average nonveteran at the Midwest Tech College of Agriculture with respect to Eighth-Quarter Grades when differences appearing in grades for the first two quarters are taken into account.

6. *Summary*

In all five comparisons of interrupted veterans with male nonveterans, the veteran subgroup was found to be superior to nonveterans in achievement relative to ability. In these five comparisons, ability was measured in terms of freshman college grades. In four of the five comparisons the difference was significant at the 1 per cent level. The differences were generally greater in comparisons involving interrupted veterans than in the comparisons involving freshman students where ability was measured by tests and high-school standing.

In all five of the studies it was noted that the interrupted veterans had lower mean scores for freshman grade than the nonveteran male students. This fact must be taken into account in interpreting the analyses of covariance results. For example, if the lower freshman grades are the result of a tendency to let down on effort in anticipation of being drafted, the effects of the interruption would be primarily a matter of depressing achievement before service rather than enhancing achievement after service.

In addition to the possibility that the freshman grades of the interrupted veterans were low because of a feeling that college work was unimportant in view of the war emergency, particularly since their college career might be interrupted before the end of the college year, there are other reasonable interpretations. One possibility is that there might have been a tendency for faculties to give higher grades in 1945, when the nonveterans were freshmen, than in the period two or three years earlier when the veterans were freshmen. Another possibility is that selective factors of some sort were operating which tended to lower the means for the interrupted veteran subgroups who returned without advanced standing gained in the various college training programs sponsored by the armed services. This seems a likely possibility, since students who were known to have had college training while in the service were excluded from the study.

No direct evidence is available bearing on the hypothesis of a letdown in anticipation of being drafted except the slight suggestion from the separate semester means at Adams that there might have been a letdown in the second term: the second-semester mean grade increased slightly over the first-semester mean grade for nonveterans, while for veterans there was a slight decrease. A similar result was not found at Stewart, however. It would appear that if there was a letdown, it affected both terms of the freshman year to about the same extent.

Evidence from Eastern City and Stewart University records shows that mean freshman average grade is rather stable from year to year, which casts doubt on the hypothesis of grades drifting upward. In addition, it has been shown that the relationships of the mean grades are the same at Midwest Tech, where the nonveterans entered in 1939, as at the other three institutions. The hypothesis of grades drifting upward over the years is not necessary to account for the differences at Adams. The percentile equivalents of semester averages were obtained, based on the entire class for that term; the percentile equivalents of the First- and Second-Semester Averages for the median interrupted veterans were about 35, while for the nonveterans the percentiles were about 55.

The most probable explanation of the lower freshman grades of the interrupted veterans is that selective factors were operating. Students with the best academic records are somewhat more likely to have been rejected in our editing procedure because they were retained in ASTP or V-12 college training programs or because they were selected for other types of military training which resulted in their being given advanced standing.

F. TRENDS IN THE COLLEGE GRADES OF VETERANS AND NONVETERANS

As has previously been mentioned, it is possible that veteran students are initially handicapped because of such factors as forgetting and deterioration of study skills, which result in temporary underachievement. Another possibility is that the veterans return to college fired with enthusiasm and good intentions, which lead to temporary overachievement. It is also, of course, possible that both these effects occur, in the same student or in different students, and that they tend to cancel each other. It was accordingly thought worth while to examine the trends in grades of veteran and nonveteran students.

Average grades by terms were obtained for six groups, including the groups at Adams and Stewart which involved interrupted veterans. The average grades earned in successive terms by veteran and nonveteran male students are shown in Figures 3 and 4. It will be noted that the entire scale of grades has been shown on the graphs for all universities except Adams. This fact should be kept in mind in interpreting the graphs, since the variation among average grades for large groups of students is not likely to be great relative to the complete scale of grades in use.

The results for interrupted veterans at Stewart and Adams, shown graphically in Figure 3, have already been described in the preceding discussions of the academic data. At Adams, the veterans do not show the slight increase in second-semester average grade which is present for the nonveterans, which suggests there may have been some letdown in effort in anticipation of being drafted. At Stewart, however, no such trend is found. Apparently any reduction in effort resulting from the imminence of withdrawing from college for military service had about the same effect on both semesters. At Adams, the line showing average grades for the semesters following return from war service is almost exactly parallel to the line for nonveterans. At Stewart there is some suggestion that the veterans did relatively better in the third term than in the fourth. The absence of any marked trend in the second-year mean grades of interrupted veterans suggests that forgetting or deterioration of study skills was not an important determiner of postwar achievement. If such deterioration did occur, it was apparently more than counter-

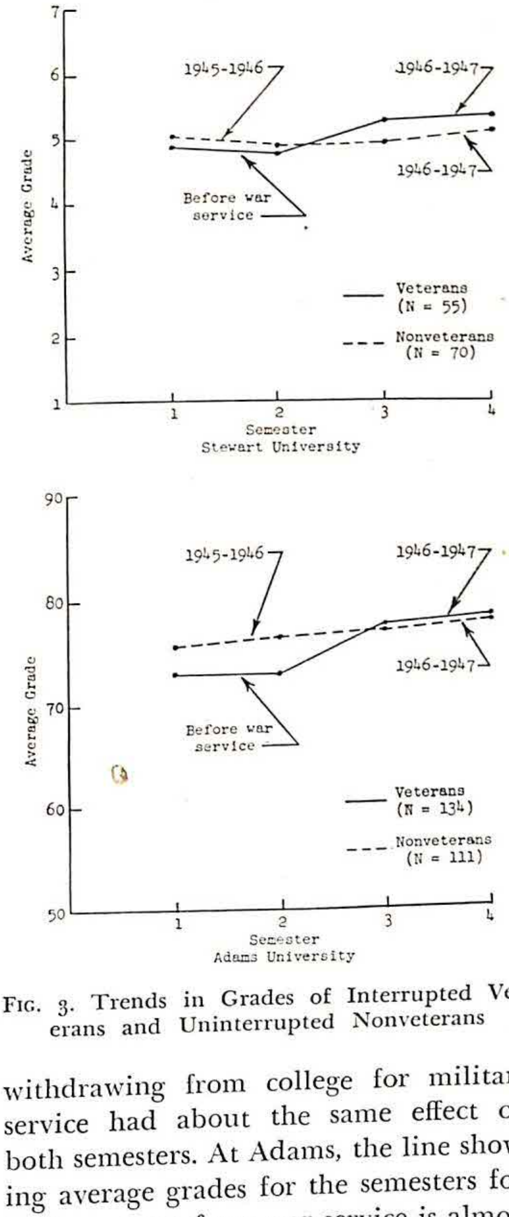


FIG. 3. Trends in Grades of Interrupted Veterans and Uninterrupted Nonveterans

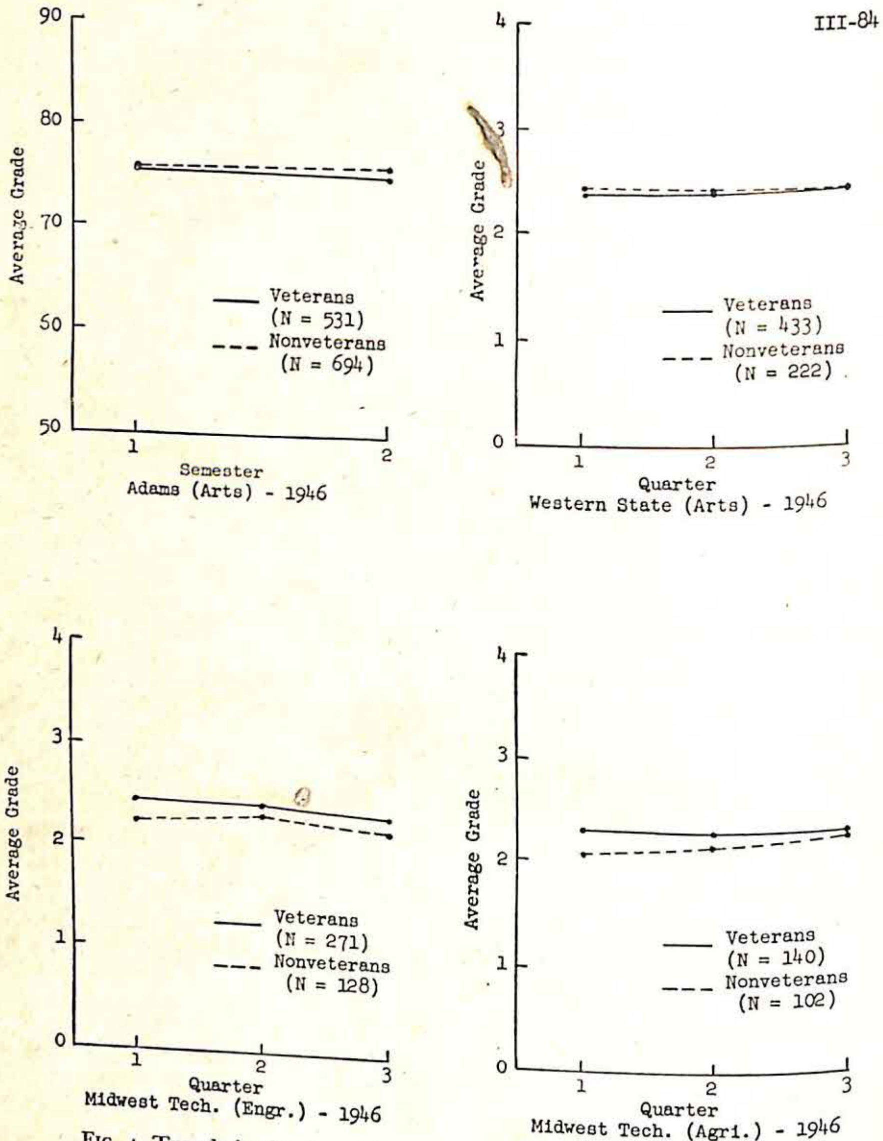


FIG. 4. Trends in Grades of Veteran and Nonveteran Male Students

balanced by influences favorable to success.

Mean term grades were computed for groups of veteran and nonveteran males who entered as freshmen in the fall of 1946 at three institutions—Adams, Western State, and Midwest Tech; at the latter institution, such data were available for both engineering and agriculture students. These means are shown graphically in Figure 4. The lines for veterans and nonveterans are nearly parallel in all four cases. At Midwest Tech there is a slight tendency for the difference between veterans and nonveterans to be greatest in the first of the three terms. There is again no consistent evidence that veterans are more handicapped at the start because of such factors as forgetting; if anything, the tendency is for veterans to do relatively better in the first term after war service than in subsequent terms.

III. SUMMARY AND INTERPRETATION

The academic success of veterans in college was early recognized as an important subject for investigation because of its relevance to long-range formulations of educational and public policy. It was clear that the performance of veterans had an important bearing on two major issues in higher education: who should go to college? and at what age should the typical student enter college? It was also clear that the performance of veterans in college was relevant to the problem of predicting college success, a matter which has received considerable attention from psychologists, especially during the past thirty years.

A major purpose of the present study was to provide a reasonably clear answer to the question: how did veterans and nonveterans differ with respect to academic success? On examination of this question, it became evident that a number of specific steps must be taken in order to minimize influences which might obscure the differences and thus present a misleading estimate of their importance. In particular, the following procedures were carried out:

1. Recognizing the diversity of American colleges and universities, studies were made in cooperation with 16 colleges and universities so chosen as to include private colleges, state universities, and municipal universities; coeducational and men's colleges; large universities and relatively small colleges; colleges with great financial resources and less wealthy institutions; and colleges located in large cities as well as colleges in small towns.

2. The crucial comparisons in this study were between *male* veterans and *male* nonveterans who were enrolled in the *same* division (e.g., liberal arts or engineering), in the *same* class (e.g., freshman or sophomore), of the *same* college or university.

3. Differences between veterans and nonveterans in aptitude for college, as measured by conventional predictors of college success, were controlled by the use of analysis of covariance.

4. Those veterans who had received a substantial amount of college training in the V-12 program or in the Army Specialized Training Program were excluded from the comparisons.

Separate comparisons were made for veteran and nonveteran students in each of 25 separate groups, each of which was homogeneous with respect to institution, division within the institution, and academic status of the students. In effect, in each college group veteran students were compared with nonveterans *of the same ability*. The estimates of ability which were employed were based, in most instances, on a test or tests of scholastic aptitude and achievement, or on a measure of high-school success used in combination with a test or tests of aptitude and achievement.

What, then, were the findings with respect to veteran-nonveteran differences in Adjusted Average Grade? In Figure 5, 25 separate answers are provided, one for each group studied. Let us consider first the results for the 20 groups containing freshman students.

For ease of interpretation, the results are presented in terms of the percentage of veterans excelling the average nonveteran in freshman average grade adjusted in such a manner that any ability differences between veterans and nonveterans have been cancelled. If veterans

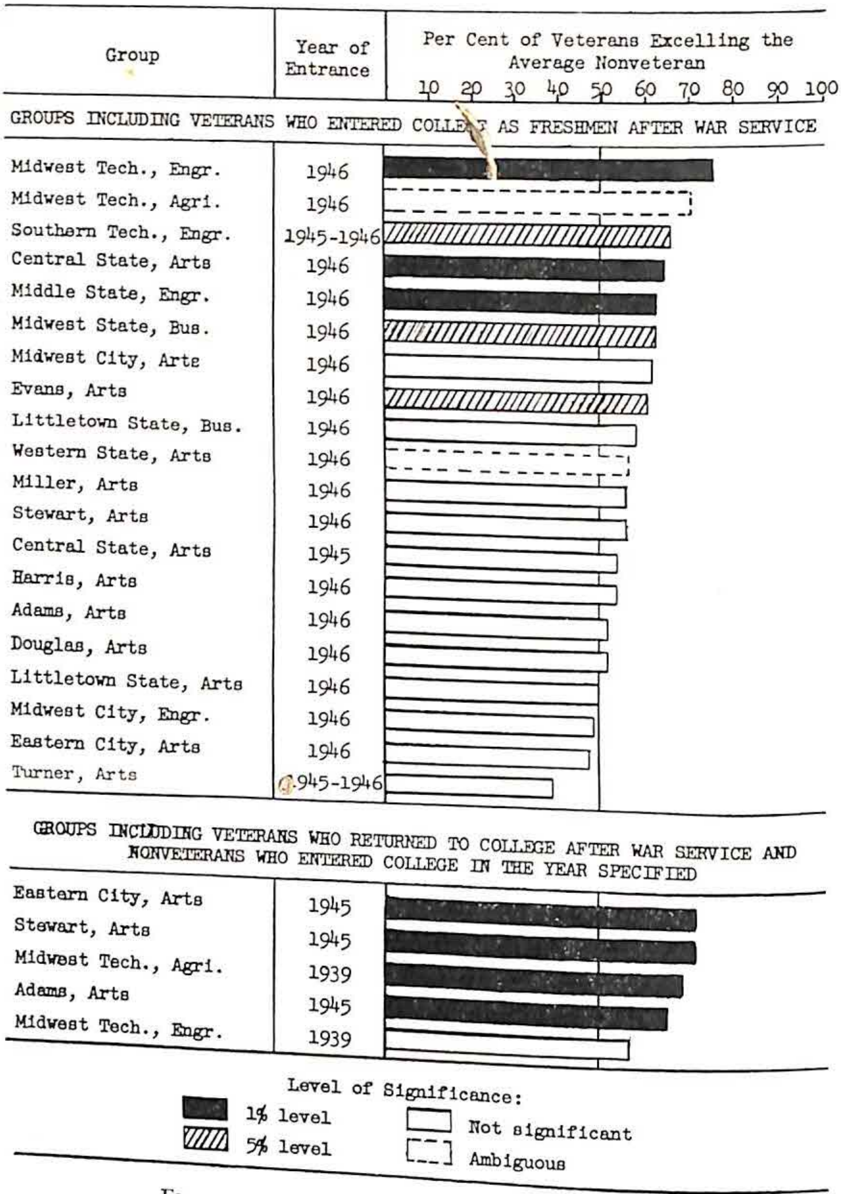


FIG. 5. Per Cent of Veterans Excelling the Average Nonveteran with Respect to Adjusted Average Grade in Each of Twenty-Five College Groups

and nonveterans were exactly alike, the two adjusted-grade distributions and the mean adjusted grades would coincide, and 50 per cent of the veterans would be found to excel the average nonveteran. If veteran students are slightly superior, the two distributions will not coincide but will overlap, and that portion of the veterans' distribution which falls to the right of the mean for nonveterans will exceed 50 per cent.

Referring to Figure 5 we may note the two most extreme cases: 76 per cent of the veterans excelled the average nonveteran in adjusted grade in the engineering school of Midwest Tech (a midwestern land-grant college), while at the opposite extreme only 39 per cent of the veterans excelled the average nonveteran at Turner (a private coeducational university). In the latter case the veterans were inferior to the nonveterans.

Veteran superiority in grades relative to ability is then not a universal tendency. What about the over-all results? It will be observed that in 16 of the 20 comparisons involving freshman students the veteran subgroup was superior; one would expect by chance to find 16 out of 20 differences in one direction less than 5 times in a hundred trials. This result in itself, being significant at the 5 per cent level, may be considered as moderately convincing evidence of superiority of veteran students in academic work when ability differences are kept constant. In the remaining four comparisons the nonveterans were superior (although for liberal arts students at Littletown State the superiority of the nonveterans is too small to be apparent in Figure 5).

Are the differences between veterans and nonveterans statistically significant in each of the 20 college groups now being discussed? Evidence on this question is shown by the shading of the bars in Figure 5. The black bars indicate a highly significant difference (the 1 per cent level of confidence), the diagonally shaded bars indicate significance at the 5 per cent level, and the white bars indicate results which are not significant at the 5 per cent level. Bars outlined with broken lines indicate that the results of the significance test were ambiguous (because of technical considerations related to the fact that errors of estimate are greater for one subgroup than for the other).

In three of the comparisons, those for engineering students at Midwest Tech and Middle State and for liberal arts students at Central State who entered in 1946, the veterans are significantly superior (at the 1 per cent level) to nonveterans. In three additional comparisons the differences are significant at the 5 per cent level. The remaining ten comparisons which showed veterans to be superior were either ambiguous or not significant (at the 5 per cent level). In none of the four cases where nonveterans were superior was the difference significant.

On the whole, therefore, it must be concluded from the studies of freshman students that there *is* a tendency for veterans to achieve higher grades in relation to ability than do nonveteran students, but the magnitude of the difference is small. In the most extreme case, the advantage of the veterans would on the average amount to no more than the difference between a "C" and a "C+." In other institutions the difference was much smaller, and, as we have seen, the direction of the difference is even reversed in some colleges.

Now let us consider the remaining five comparisons—those involving "interrupted" veterans. Substantially all the veteran students in these five comparisons were those who had completed at least two quarters of freshman work as ordinary civilians, who then entered military service, and who returned to college after discharge and completed at least an additional two quarters of academic work. The nonveterans with whom the interrupted veterans were compared were ordinary civilian students who completed without interruption the same amount of academic work as the veterans. In three of the college groups these nonveterans were postwar students who entered college as freshmen in 1945, but at Midwest Tech they were prewar students who had entered in 1939. In all five of these groups, college grades earned early in a student's academic career were taken as the measure of his ability; those earned later in his college career were used as the measure of his academic success. For the veterans, of course, a considerable amount of time elapsed between their early and later academic work.

In all five groups involving interrupted veterans, veteran students were superior to nonveterans of equal ability. In four of the five comparisons, the difference in adjusted grade was significant at the 1 per cent level. The proportion of veterans who excelled the average nonveteran ranged from 72 per cent at Eastern City and Stewart to 57 per cent at the Midwest Tech engineering school. The evidence from these studies of interrupted students thus strongly supports the hypothesis that veterans do excel nonveterans of equal ability with respect to achievement in college.

The interpretation of the differences found between veterans and nonveterans in these five groups is particularly difficult. At first glance, it would appear that the veterans had gained greater maturity and a capacity for more intensive and prolonged effort during their absence from the campus; accordingly, they showed a marked gain over their initial performance—in contrast to the relative stability of performance shown by a nonveteran group whose careers were not interrupted. Such an interpretation would overlook two important complications: first, the possibility that shifts in grading standards over a period of time might influence the results; and second, the fact that in all five groups the veterans earned *lower* freshman grades than did their nonveteran controls. The hypothesis that grading standards shifted is weakened by the fact that in three of the groups the veterans were compared with a postwar nonveteran group (students entering in 1945) while in the other two the nonveteran comparison group was a prewar group (students entering in 1939). The advantage of the nonveterans in their *freshman* grades constitutes a more serious problem. Among the hypotheses which may be advanced to account for this finding, two appear to be most plausible: first, the members of the veteran group, realizing that their academic careers were likely to be interrupted by war service, may have slackened their efforts; and second, the necessary exclusion from the comparisons of veterans who continued their academic careers in V-12 or ASTP may have removed a disproportionate number of superior students, and thus lowered the average in the remaining group. With these limitations in mind, it must be recognized that the findings are not as conclusive as the significance tests might indicate. The findings, in spite of the limitations, may properly be considered to favor the veteran group.

In addition to the comparisons of male veterans and male nonveterans, study was made of a number of problems which were considered pertinent to the interpretation of the findings or which were by-products of the plan of analysis followed in making the basic comparisons. These problems were studied in varying numbers of groups; no attempt was made, however, to exhaust the possibilities of the available raw data in seeking answers to them.

One question, studied in three groups, was the possible significance of academic load, as measured by the number of credit hours carried. In all three instances, it was found that veterans tended to take a slightly lighter load. This might merely be a reflection of the fact that veterans were generally not required to take physical education or military science. It was also found that correlations between work load and measures of ability and college achievement are consistently positive—students who take the heavier load tend to get the higher grades in college. This tendency was consistently found both for the veteran subgroups and the nonveteran subgroups. This finding tends to discount the hypothesis that veterans did better because of a reduced course load, although it does not provide an adequate basis for rejecting it.

It is important to study the relationship between the time when aptitude tests were taken by veterans and the scores which they earned. In two institutions where the College Entrance Examination Board Scholastic Aptitude Test was used, some veterans had applied for admission and were tested before leaving for war service and some were tested after discharge. This variable, date of testing, was correlated with all the measures of ability and college achievement. The correlations tended to be quite low, indicating that date of testing, within the range encountered in these studies, is of little importance. This finding is reassuring in the interpretation of the results, since in most of the groups the veterans were tested after war service, several years after high-school graduation, while the nonveterans were tested a few months after graduation from high school.

It is pertinent to inquire whether any difference in outcome of analyses of covariance is apparent in the groups for which both high-school standing and various test scores were used as compared with those groups for which only test scores were used. As it happened, there were 10 groups of each type among the 20 groups of entering freshmen. The general outcome of this comparison is shown in Figure 6. The median value of the percentage of veterans excelling the average nonveteran is 56 in both sets. However, the range in percentage of veterans excelling the average nonveteran

Per Cent of Veterans Excelling the Average Nonveteran	Predictors Used	
	High School Standing in Combination with a Test or Tests	A Test or Combination of Tests without High School Standing
78 or higher		
73 to 77	•	
68 to 72	×	
63 to 67	••	••
58 to 62		•oo
53 to 57	ooo×	o
48 to 52	oo	ooo
43 to 47		
38 to 42		o
33 to 37		
28 to 32		
23 to 27		
22 or lower		

• Difference significant at or beyond 5% level

o Difference not significant

× Outcome of analysis ambiguous

FIG. 6. Distribution of Per Cent of Veterans Excelling the Average Nonveteran with Respect to Adjusted Average Grade for the Ten Groups in Which High-School Standing Was Used in Combination with a Test or Tests and for the Ten Groups in Which Tests Only Were Used

is 48 to 76 in the groups where high-school standing was used, and the range is 39 to 66 in the groups where tests alone were used. This suggests that the inclusion of high-school standing as a predictor is associated with findings which are slightly more favorable to the veteran group.

Finally, attention was given to possible trends in grades which veteran students earned in successive terms after war service. Examination of the mean grades of veterans and nonveterans shows no evidence that veterans were seriously handicapped at the beginning of the college work because of forgetting or deterioration of study skills, nor that there was a period of initial enthusiasm which led to marked temporary overachievement. In four comparisons of this type, no clear indication of a consistent shift in differences of grades was found; if forgetting does play a role in grades earned by veterans immediately after their entrance to college, it is apparently counterbalanced by other factors such as a more enthusiastic approach to the study situation.

Accepted for publication, February 28, 1952.

Vol. 66, No. 16 Whole No. 348, 1952

Psychological Monographs: General and Applied

The Primary Personal Values Measured by the Allport-Vernon Test, "A Study of Values"*

Hubert E. Brogden
Personnel Research Section, AGO, Department of the Army

I. INTRODUCTION

Since its introduction, the Allport-Vernon test, *A Study of Values,* has attracted widespread interest among psychologists. A number of authors have reported the relations of the value scores to each other and to other variables. Few, if any, have concerned themselves with the factors accounting for the internal relations among the test items. Many psychologists would consider this latter problem basic to the determination of the relationship between values and other variables. If the measures employed overlap or intercorrelate considerably, confusion in the interpretation of values in relation to other measures appears likely. It is the purpose of the present investigation to determine the primary traits necessary to account for the intercorrelations of 60 items of this test. Such an analysis should add to our understanding of personal values in general and have some significance for the study of interest and personality traits.

The manual accompanying the first edition of the test reported high correlation between several of the value scores, thus suggesting that these measures were probably not independent and that they presumably measure to a considerable extent common value tendencies. Since that time other investigators have verified the correlation originally reported between these measures. Duffy and Crissy (2) reported the isolation of three factors from the correlations between the various value scores. However, the first and most extensive factor study of the Allport-Vernon test was that of Lurie (5). Lurie, being somewhat critical of the multiple-choice structure of the Allport-Vernon test items, constructed a similar test whose constituent items gave an indication of degree of evaluation without requiring a choice between several objects of evaluation. Twenty-four variables were employed in the analysis, each variable being the sum of several items. Of the seven factors necessary to account for the intercorrelations of these 24 variables, four were regarded by Lurie as meaningful. The six original evaluative types, namely the aesthetic, religious, social, economic, political, and theoretic, were reduced to four by the disappearance of the aesthetic evaluative dimension and the combination of the political and economic types.

II. DESCRIPTION OF VARIABLES

Basic to the construction of the Allport-Vernon test is Spranger's theory that men are best known through a study of their subjective values. Spranger proposed a sixfold classification of such values as follows: the *theoretical,* the *economic,* the *social,* the *political,* and the *religious.*

---

* The opinions expressed are those of the author, and do not necessarily reflect the official viewpoint of the Department of the Army.

The test of values developed by Allport and Vernon (1) consists of a number of questions based upon a variety of familiar situations, to which 2 alternative answers (in Part I) and 4 alternative answers (in Part II) are provided. There are 120 alternative answers in all, 20 referring to each of the six values. Subjects are instructed to mark the alternative answers in Part I according to degree of preference and to rank answers in Part II in order of their choices.

Sample questions are given below with responses of a hypothetical subject.

Part I. Question 1.

The main object of scientific research should be the discovery of pure truth rather than its practical applications.

	(*a*)	(*b*)
(*a*) Yes (*b*) No	3	0

Part II. Question 14. (Women have an alternative form of this question.)

If you should marry (or are married) do you prefer a wife who—

4 *a.* can achieve social prestige, commanding admiration from others;
1 *b.* likes to stay at home and keep house;
2 *c.* is fundamentally spiritual in her attitude toward life;
3 *d.* is gifted along artistic lines.

It can be seen from the nature of the items that the test scores give the *relative* prominence of the several values types.

Split-half reliabilities reported by the authors (7) range from .49 for the "social" scale to .84 for the "aesthetic" and "religious" scales. Test-retest reliabilities range from .39 for the "social" scale to .84 for the "aesthetic" scale.

Although it was desired that the present study be as comprehensive as possible, 30 possible variables were excluded from the analysis in order that computational labor might be kept within reasonable bounds. In the first section of 30 items, where the subject is instructed to choose between 2 alternative statements or to accept or reject a single statement, each item was considered a single variable. In the second section of 4 alternative items, each alternative can be considered a separate variable, making a possible total of 60 variables for this section. In this latter section, the 4 alternatives of each item have forced negative intercorrelations. In order to reduce such forced negative correlation to a minimum, the first 2 alternatives of each 4 alternative items were discarded. The variables for the present analysis were thus confined to the 30 items of Section 1, and to 30 of the 60 item alternatives in Section 2 of the Allport-Vernon test.

The variables are listed with their numbering in the present analysis and with an indication of the value they were designated as measuring in the scoring key provided with the test of values. The letters A, R, E, P, S, and T are used to show the manner of scoring and are abbreviations for, respectively, *aesthetic, religious, economic, political, social,* or *theoretic* personal values. Since the last 30 variables are alternatives of four-choice items which were to be ranked in order of preference, all four alternatives must be shown in order to describe fully the variable. For example, one item in Part II calls for the ranking of four persons according to the subject's interest in them. The ranking assigned one of these four persons consitutes a variable of the factor analysis. It is listed as follows: "Florence Nightingale interests and attracts him more than Napoleon, Henry Ford, or Charles Darwin"—this is listed as our item No. 55, below. The letters "S vs. P, E, T" following item

No. 55 indicate that choice of Florence Nightingale was scored as social, while the three alternative choices were scored, respectively, as political, economic, or theoretic.

It was convenient in the factor analysis to reverse the sign of certain of the variables. When this was done, the statement of the variable was altered to conform with the sign reversal. The reversed statement of the sample item given above would read: "Napoleon, Henry Ford, or Charles Darwin interests or attracts him more than Florence Nightingale."

The variables of the analysis are listed below:

1. He believes that the main object of scientific research should be the discovery of pure truth rather than its practical applications. (T vs. E[1])

2. He believes that it is justifiable for the great artists such as Beethoven to be selfish and negligent of the feelings of others. (A vs. S)

3. He does not believe that because of the aggressive and self-assertive nature of man, the abolition of war is an illusory ideal. (R vs. P)

4. If he were a university professor and had the necessary ability, he would prefer to teach poetry rather than chemistry and physics. (A vs. T)

5. He would prefer to teach economics rather than law. (E vs. P)

6. He considers unselfishness and sympathy more desirable character traits than high ideals and reverence. (S vs. R)

7. In a paper such as the *New York Sunday Times,* he is more interested in the section on picture galleries than in the real-estate section and the account of the stock market. (A vs. E)

8. He does not think that a person who analyzes his emotions is apt to be less sincere in his feeling than one who is not so reflective. (T vs. S)

9. He should prefer to read the headlines "Protestant Leaders To Consult On Reconciliation" rather than "Great Improvements In Market Conditions." (R vs. E)

10. He would prefer to read the story "Physicists Make Important Discoveries Concerning The Structure Of Atoms" rather than "Threat To Constitutional Government In Germany By A Dictator." (T vs. P)

11. When he visits a cathedral, he is more impressed by the architectural features and stained glass than by a pervading sense of reverence and worship. (A vs. R)

12. He does not believe that contemporary charitable policies should be curtailed because they tend to undermine individual initiative. (S vs. P)

13. He believes that the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first cause, cosmic purpose, or God behind it. (T vs. R)

14. In his opinion, general progress has been advanced more by the freeing of slaves with enhancement of the value placed on individual life rather than by the discovery of the steam engine with the consequent industrialization and economic rivalry in American and European countries. (S vs. E)

15. If he had the opportunity, and if nothing of the kind existed in the community or college where he lived, he would prefer to found a classical orchestra rather than a debating society. (A vs. P)

16. At an exposition he prefers to see scientific apparatus or chemical products rather than automobiles. (T vs. E)

17. He would prefer to hear a series of popular lectures on the contemporary painter rather than the progress and needs of social-service work in the cities of his part of the country. (A vs. S)

18. He would prefer to hear a series of lectures on the comparative development of the great religious faiths rather than the comparative merits of the forms of government in Britain and in the United States. (R vs. P)

19. If he had some time to spend in a waiting room, and there were only two magazines to choose from, he would read *Arts and Decorations* rather than *The Scientific American.* (A vs. T)

20. He would encourage his children to have vocational training rather than to make several teams. (E vs. P)

21. He feels that the aim of the churches at the present time should be to bring out charitable tendencies, and to urge people to think more of the good of others, rather than to convey spiritual worship and a sense of communion with the highest. (S vs. R)

22. He feels that our modern industrial and scientific developments are not signs of a greater degree of civilization and culture than those attained by any previous race, such as the Greeks. (A vs. E)

---

[1] The letters A, R, E, P, S, and T are employed to indicate that the item is, in the Allport-Vernon test, scored, respectively, as a measure of Aesthetic, Religious, Economic, Political, Social, or Theoretic value.

23. If he were waiting in a strange living room, he would be more apt to judge the owner's knowledge and intellectuality as shown by the books in his bookcases than his friendliness and hospitality as shown by the photographs on the walls and the general comforts of the room. (T vs. S)

24. He feels that the world would not be a much better place if we took to heart the teaching, "Lay not up for yourselves treasures upon earth . . . but lay up for yourselves treasures in heaven, where neither moth nor rust doth corrupt, and where thieves do not break through nor steal." (E vs. R)

25. He is more interested in reading accounts of the lives and works of such men as Aristotle, Plato, and Socrates than of Alexander, Julius Caesar, and Charlemagne. (T vs. P)

26. Taking the Bible as a whole, he would prefer to regard it from the point of view of its beautiful mythology and literary style rather than as a spiritual revelation. (A vs. R)

27. He does not believe that since the class or social status to which a man belongs depends mainly upon his push and ability, it is just that a small proportion of the population should be very rich. (S vs. P)

28. If he were to choose a topic on which to write an essay, he would choose the defects of our present educational systems rather than the role of church-going in religion. (T vs. R)

29. If he were to choose a topic on which to write an essay, he would prefer the personality of some close friend of his rather than the necessities of life, luxuries, and savings. (S vs. E)

30. When witnessing a gorgeous ceremony (ecclesiastical or academic, induction into office, etc.) he is more impressed by the color and pageantry of the occasion itself rather than by the unified idea or institution which the group represents. (A vs. P)

31. He believes that good government should aim at more aid for the poor, sick, and old; rather than the development of manufacturing and trade, introduction of more ethical principles into its policies and diplomacy, or the establishing of a position of prestige and respect among nations. (S vs. E, P, R)

32. He believes that good government should aim at more aid for the poor, sick, and old, the introduction of more ethical principles into its policies and diplomacy, or the establishment of a position of prestige and respect among nations; rather than the development of manufacturing and trade. (S, P, R vs. E)

33. He believes that a man who works in business for his living all the week can best spend Sunday in trying to educate himself by reading serious books; rather than by trying to win at golf or racing, going to an orchestral concert, or hearing a really good sermon. (T vs. P, R, A)

34. He believes that a man who works in business all week can best spend Sunday in trying to educate himself by reading serious books, going to an orchestral concert, or hearing a really good sermon; rather than by trying to win at golf or racing. (R, A, T vs. P)

35. If he could influence the educational policies of the public schools of some city, he would undertake to promote the study and the performance of drama; rather than to develop cooperativeness and the spirit of service, to provide additional laboratory facilities, or to promote school savings banks for education in thrift. (A vs. T, E, S)

36. If he could influence the educational policies of the public schools, he would undertake to promote the study and the performance of the drama, to provide additional laboratory facilities, to promote school savings banks for education in thrift; rather than to develop cooperativeness and the spirit of service. (A, T, E vs. S)

37. He would prefer a friend of his own sex who is seriously interested in thinking out his atttiude toward life as a whole, one who possesses qualities of leadership and organizing ability, one who shows refinement and emotional sensitivity; rather than one who is efficient, industrious, and of a practical turn of mind. (S, P, R vs. E)

38. He would prefer a friend of his own sex who is seriously interested in thinking out his attitude toward life as a whole; rather than one who is efficient, industrious, and of a practical turn of mind, one who possesses qualities of leadership and organizing ability, or one who shows refinement and emotional sensitivity. (R vs. S, P, E)

39. If he lived in a small town and had more than enough income for his needs, he would prefer to use it to endow the church to which he belonged, give it to a university for the development of scientific research, devote it to hospitals; rather than apply it productively to industrial development. (T, S, R vs. E)

40. If he lived in a small town and had more than enough income for his needs, he would prefer to apply it productively to industrial development, give it to a university for the development of scientific research, devote it to the hospitals; rather than use it to endow the church to which he belongs. (T, S, E vs. R)

41. When he goes to the theatre, he enjoys most, as a rule, the ballet or similar imaginative performances, plays with a theme of human suffering and love, problem plays that argue consistently for some point of view; rather than plays that treat the lives of great men. (T, A, S vs. P)

42. When he goes to the theatre, he enjoys most, as a rule, ballet or similar imaginative performances; rather than plays that treat the lives of great men, plays with a theme of human suffering and love, or problem plays that argue consistently for some point of view. (A vs. T, S, P)

43. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician; rather than a sales manager, clergyman, or politician. (T vs. E, R, P)

44. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician, clergyman, politician; rather than a sales manager. (T, R, P vs. E)

45. If he had unlimited leisure and money, he would prefer to make a collection of fine sculptures or paintings; rather than establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance. (A vs. S, P, E)

46. If he had unlimited leisure and money, he would prefer to establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient; rather than make a collection of fine sculptures or paintings, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance. (S vs. A, P, E)

47. At an evening discussion with intimate friends of his own sex, he is most interested when they talk about philosophy and psychology, literature, socialism and social amelioration; rather than the meaning of life. (T, A, S vs. R)

48. At an evening discussion with intimate friends of his own sex, he is most interested when they talk about philosophy and psychology; rather than the meaning of life, literature, or socialism and social amelioration. (T vs. A, S, R)

49. During his next summer vacation he would prefer to write and publish an original biological essay or article; rather than stay in some secluded part of the country where he could appreciate fine scenery, go in for a local tennis or other athletic tournament, or get experience in some new line of business. (T vs. E, A, P)

50. During his next summer vacation he would prefer to stay in some secluded part of the country where he could appreciate fine scenery; rather than write and publish an original biological essay or article, go in for a local tennis or other athletic tournament, or get experience in some new line of business. (A vs. T, E, P)

51. Great exploits and adventures of discovery such as Lindbergh's and Byrd's seem significant because they add to our knowledge of mechanics, geography, meteorology, etc., because they weld human interests and international feelings throughout the world, or because they contribute to the ultimate revelation of the meaning of the universe; rather than because they represent conquests by man over the difficult forces of nature. (T, S, R vs. P)

52. Great exploits and adventures of discovery such as Lindbergh's and Byrd's seem significant to him because they add to our knowledge of mechanics, geography, meteorology, etc.; rather than because they represent conquests by man over the difficult forces of nature, because they weld human interests and international feelings throughout the world, or because they contribute to the ultimate revelation of the meaning of the universe. (T vs. S, R, P)

53. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's ideals of beauty, one's business organizations and associates, society as a whole; rather than toward one's religious faith. (A, E, S vs. R)

54. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's ideals of beauty; rather than one's religious faith, one's business organizations and associates, or society as a whole. (A vs. E, S, R)

55. Florence Nightingale interests and attracts him more than Napoleon, Henry Ford, or Charles Darwin. (S vs. P, E, T)

56. Florence Nightingale, Henry Ford, or Charles Darwin interests and attracts him more than Napoleon. (S, E, T vs. P)

57. He would prefer, if he were to marry, a wife who would like to stay at home and keep house, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who can achieve social prestige commanding admiration from others. (A, S, R vs. P)

58. He would prefer, if he were to marry, a wife who could achieve social prestige commanding admiration from others, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who likes to stay at home and keep house. (A, R, P vs. S)

59. He would like to think of Leonardo da Vinci's "The Last Supper" as one of the most priceless and irreplaceable pictures ever painted, in relation to Leonardo's versatility and its place in history, as a masterpiece of design; rather than as one expressing the highest spiritual tendencies and emotions. (E, T, A vs. R)

60. He would like to think of Leonardo da Vinci's picture "The Last Supper" as expressing the highest spiritual tendencies and emotions, in

relation to Leonardo's versatility and its place in history, as a masterpiece of design; rather than as one of the most priceless and irreplaceable pictures ever painted. (R, T, A vs. E)

III. POPULATION

The subjects of the present study were 200 male students of the University of Illinois. The limitation to one sex was imposed because of the probability that males would differ from females not only in the magnitude of their evaluative tendencies but also in the interrelations of such tendencies. The test papers used were those of students who had been tested in 1936 and 1937 as a part of a regular class exercise during a course in social psychology.

IV. PROCEDURE

Tetrachoric intercorrelations between each of the 60 variables were computed and the resulting matrix subjected to a centroid analysis. After the extraction of eleven centroid factors, the residuals were symmetrically distributed and had a standard deviation of .08. This may be compared with the standard error of .11 of a tetrachoric correlation of zero with a median split of both variables. Rotation of the centroid factors to an oblique simple structure was next accomplished.

The intercorrelations of the primary traits were then determined. Since a number of these correlations appeared significantly high, the matrix formed by variables involved in such significant correlations was factored and the resulting centroid factors rotated orthogonally into a simple structure. Factors III, IX, and X were excluded from this analysis for the reason just mentioned, that is, their correlations with other primary traits were quite low. Factor XI was also not included since it is residual in nature. Loadings of the 60 original variables on these second-order factors were then calculated.

In arriving at the general factor solution presented as Table 6, a number of alternatives were investigated and discarded as being less meaningful than the one here presented. Throughout the investigation of these alternative solutions, the factor pattern showed little change with respect to variables with loadings above .4. Variables with loadings between .2 and .4 could be made to vary above and below the arbitrary level of significance, namely .30. This lack of rigidity in the simple structure is evident in the literature on factor analyses in nearly all fields other than abilities. Its effect will probably be greatest in loadings on the second-order factors.

Since the reader may feel, in considering some of the following discussion of the individual primary traits, that an undue amount of detailed discussion is devoted to the relation between individual variables and the primary traits of this and Lurie's analysis, the reasons for this emphasis will be indicated. The author believes that the identification of a primary trait is an extremely tentative process, at least with limited information at present available, and even more so that the identification of two primary traits emerging from different analyses as identical must depend upon something more than similarity between names attached to measures subjected to factor analysis. It has been shown a number of times that questionnaires purporting to measure the same thing intercorrelate negligibly. A detailed comparison of variables with high loadings on the factors in question is, consequently, desirable before statements regarding identity or similarity of factors are made.

From each of the columns of Table 6, the primary-trait matrix, a list of those variables with loadings greater than .30 has been prepared. *These are the basic definitions of the nature of each factor.* In the present instance, with so little known concerning primary traits in evaluative tendencies, most readers will prefer to arrive at their own interpretations of the individual factors by examining these lists. The writer has, however, attempted to assemble all evidence which may be of aid in this interpretation and to develop those hypotheses which seem to him most fruitful. Because of the difficulty in selecting a single word which conveys exactly the meanings necessary to define the configuration of loadings, the significance of the names assigned these factors should be minimized. Names will be employed as symbols referring to the list of traits or to hypotheses concerning these lists.

V. RESULTS

A. General

The centroid matrix is presented as Table I; the transformation from the centroid to the primary-trait matrix as Table 2; the cosines between the perpendiculars to the hyperplanes defining the primary traits as Table 3; the correlations between the factors as Table 4; the loadings of the first order on the second order factors as Table 5; and the loadings of the variables on the primary traits as Table 6. Table 6 includes the loadings of the variables on the second-order factors, *A, B,* and *C*. The variables of the analysis are designated by arabic numbering, rotated factors by roman numerals, and the centroid factors by roman numerals with the prime symbol.

B. The First-Order Factors

Factor I: General Aesthetic Interest

Variable 17. Loading, .66. He would prefer to hear a series of popular lectures on the contemporary painter rather than on the progress and needs of social-service work in the cities of his part of the country.

Variable 45. Loading, .66. If he had unlimited leisure and money, he would prefer to make a collection of fine sculptures or paintings; rather than to establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

Variable 18. Loading, .60. He would prefer to hear a series of lectures on the comparative development of the great religious faiths rather than on the comparative merits of the forms of government in Britain and in the United States.

Variable 36. Loading, .54. If he could influence the educational policies of the public schools, he would undertake to promote the study and the performance of the drama, to provide additional laboratory facilities, to promote school savings banks for education in thrift, rather than to develop cooperativeness and the spirit of service.

Variable 35. Loading, .50. If he could influence the educational policies of the public schools of some city, he would undertake to promote the study and the performance of drama; rather than to develop cooperativeness and the spirit of service, to provide additional laboratory facilities, or to promote school savings banks for education in thrift.

Variable 50. Loading, .49. During his next summer vacation he would prefer to stay in some secluded part of the country where he could appreciate fine scenery; rather than to write and publish an original biological essay or article, go in for a local tennis or other athletic tournament, or get experience in some new line of business.

Variable 19. Loading, .49. If he had some time to spend in a waiting room and there were only two magazines to choose from, he would read *Arts and Decorations* rather than *The Scientific American.*

Variable 29. Loading, .42. If he were to choose a topic on which to write an essay, he would prefer the personality of some close friend of his rather than the necessities of life, luxuries, and savings.

Variable 57. Loading, .42. He would prefer, if he were to marry, a wife who would like to stay at home and keep house, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who can achieve social prestige commanding admiration from others.

Variable 42. Loading, .36. When he goes to the theatre, he enjoys most, as a rule, ballet or similar imaginative performances; rather than plays that treat the lives of great men, plays with a theme of human suffering and love, or problem plays that argue consistently for some point of view.

Variable 4. Loading, .33. If he were a university professor and had the necessary ability, he would prefer to teach poetry rather than chemistry and physics.

Variable 41. Loading, .33. When he goes to the theatre, he enjoys most, as a rule, the ballet or similar imaginative performances, plays with a theme of human suffering and love, problem plays that argue consistently for some point of view; rather than plays that treat the lives of great men.

Variable 54. Loading, .33. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's ideals of beauty; rather than one's religious faith, one's business organizations and associates, or society as a whole.

TABLE 1

THE CENTROID MATRIX*

Variable	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	h²**
1	097	−064	−123	067	113	107	−184	079	141	158	140	162
2	281	−135	−226	−181	221	255	−239	136	−077	051	−098	389
3	365	093	195	059	−182	061	−096	051	−407	271	−255	536
4	268	−410	−258	−161	−164	−308	−256	143	−286	080	109	640
5	114	125	117	−073	263	155	−040	−064	−134	054	−182	201
6	132	261	−262	058	−163	246	285	−175	146	−207	−339	536
7	340	−312	082	−237	−298	237	−261	−088	−135	224	055	568
8	233	056	028	−212	−180	−174	207	036	183	−079	−061	253
9	164	−474	248	−268	−317	193	−153	−065	094	−283	161	665
10	073	273	343	−074	229	222	334	−227	069	059	140	496
11	186	422	−246	−136	−200	−018	−039	120	−124	−102	241	432
12	307	−119	−065	271	−396	275	060	208	−090	−107	205	527
13	225	412	−283	254	222	026	−233	312	−181	−212	−028	645
14	243	−291	236	032	−295	105	182	316	−130	195	057	490
15	135	−066	037	−140	101	340	−183	−196	123	133	−281	353
16	446	245	153	−234	265	078	025	131	131	−225	168	527
17	517	−063	−076	−541	302	177	045	060	125	098	−267	794
18	354	−442	376	−389	092	−014	140	−067	−171	−224	−261	794
19	285	−430	−288	067	−133	−259	−089	−110	−089	134	−327	591
20	143	052	020	098	−073	141	057	093	170	074	−297	193
21	251	560	−132	−032	−408	295	139	134	256	−038	−166	780
22	358	−134	−178	083	021	138	−207	167	174	−082	054	315
23	420	089	−044	−223	052	−158	−044	081	137	−049	063	297
24	184	438	−231	−089	050	−208	−173	−262	−130	−081	199	494
25	565	053	297	198	211	−248	−095	108	−180	−027	−082	616
26	543	547	−250	037	220	097	−189	102	−101	−079	115	792
27	408	−146	−033	421	−202	053	−088	086	188	154	−117	498
28	071	483	−294	062	−157	−102	−192	−044	115	179	122	463
29	386	−094	289	−249	−015	−228	286	158	−133	−156	−186	539
30	021	152	−131	−198	−232	299	193	−161	−127	−154	039	328
31	544	−144	−089	296	−230	−088	111	−111	−089	029	067	511
32	242	−418	193	143	−364	106	176	−109	012	082	202	525
33	302	−073	128	321	374	−224	251	−253	403	092	139	723
34	348	−334	323	213	122	082	141	232	131	081	131	519
35	465	−188	−367	−281	−125	−242	−149	−238	−054	096	220	679
36	157	−249	−280	−423	198	−235	111	−313	342	207	164	735
37	531	−136	−152	226	135	−196	130	220	−105	096	−178	549
38	442	−143	123	446	145	−111	400	209	−158	−161	103	728
39	378	−275	297	092	−126	172	−182	−140	129	−196	291	553
40	402	596	−468	109	152	095	138	−144	−131	052	124	855
41	538	009	−096	181	−127	−111	144	−135	−092	049	−191	446
42	412	250	−080	−177	143	−124	121	−187	−151	371	−059	519
43	403	123	129	209	160	145	−322	−088	195	140	−012	454
44	564	−294	146	115	140	−043	−164	−017	−062	233	080	552
45	666	−282	081	−262	097	−119	−035	021	046	034	−216	673
46	434	108	446	133	−107	079	123	−103	−094	−302	129	577
47	449	020	−110	−276	−093	−142	−178	417	281	161	149	652
48	090	260	218	−197	−102	−238	−124	350	162	−226	148	466
49	563	136	156	092	182	−156	085	314	072	109	230	602
50	136	−414	−219	−272	070	122	029	−162	−017	−058	−147	384
51	235	209	358	086	−182	−183	−079	−060	188	196	−109	397
52	139	173	255	116	−165	−102	−036	−250	208	287	070	360
53	165	668	−544	110	167	169	145	−213	−093	054	131	933
54	750	−058	−132	076	250	130	118	−075	−273	160	184	822
55	317	−317	378	237	−219	−066	−155	−217	−234	−337	−034	693
56	279	090	488	333	218	−105	−199	−119	−252	−300	−179	733
57	388	−516	−146	182	111	−097	−102	−216	185	−288	−258	734
58	211	−174	222	−144	265	−228	190	220	−332	222	254	576
59	350	456	−092	−235	−119	023	138	−064	−100	171	−043	473
60	042	−169	−279	336	222	−110	080	211	207	−269	−114	462

* Decimals properly preceding each entry have been omitted.

** Communality.

TABLE 2

The Transformation Matrix*

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
I′	463	325	086	315	265	174	054	264	239	−016	255
II′	−338	−103	−194	541	332	−230	159	−263	384	−062	057
III′	−243	−037	039	−532	466	−047	510	−045	158	−215	117
IV′	−564	−137	−200	−028	207	286	−010	526	−224	126	−032
V′	013	185	268	−074	−075	−520	375	042	−289	287	−455
VI′	−158	476	−251	−061	−186	262	568	−014	−365	−477	−018
VII′	244	−658	−227	−047	−491	−173	189	526	−003	007	130
VIII′	−241	219	−312	−460	−145	−366	038	463	550	−305	−403
IX′	−027	300	−347	−215	−368	089	329	086	328	649	424
X′	025	056	301	−236	−153	−394	−043	169	−204	−318	589
XI′	−389	173	648	001	−317	409	320	238	227	−092	−047

* Decimals properly preceding each entry have been omitted.

All but four of the aesthetic variables (on the original scale of values) are significantly loaded on Factor I. Consequently, Factor I may be regarded with considerable confidence as aesthetic in nature.

Variables 2, 7, 15, and 26 purport to measure aesthetic evaluative tendencies but are not loaded on Factor I. Since the loading of Variables 2, 7, and 15 are, from Table 1, .246, .281, and .214, respectively, it would appear that their absence from the above listing can be plausibly attributed to sampling error. The loading of −.021 of Variable 26 does indicate some discrepancy between the content of Factor I and aesthetic evaluation as conceived by the authors of the test.

Variables 18 and 29 are loaded on Factor I although they apparently lack aesthetic content. Subjectively, however, they appear to reveal emotional or appreciative tendencies which might be expected in persons aesthetically inclined. Variables 36 and 57 while not primarily aesthetic have, respectively, as an alternate choice "teaching drama" and "choosing a wife gifted along artistic lines." These alternatives were assumed to be adequate explanation of their loading on Factor I.

The absence of a clearly defined aesthetic factor in Lurie's results is difficult to explain. In order that this problem may be more closely analyzed, each of the variables of Lurie's analysis which were intended as measures of an aesthetic

TABLE 3

The Cosine Matrix*

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
I	1.000	−089	036	215	−147	−165	−329	−146	−113	159	236
II		1.000	110	−121	021	180	359	−125	084	−093	−037
III			1.000	061	−005	064	−021	−107	−188	−113	000
IV				1.000	230	220	−317	−349	−027	156	033
V					1.000	033	−145	−388	068	−152	−065
VI						1.000	030	011	−062	040	160
VII							1.000	166	051	−097	010
VIII								1.000	131	−051	018
IX									1.000	090	033
X										1.000	057
XI											1.000

* Decimals properly preceding each entry have been omitted.

TABLE 4

THE INTERCORRELATIONS OF THE PRIMARY FACTORS*

	II	III	IV	V	VI	VII	VIII	IX	X	XI
I	−116	−001	−186	233	277	312	088	120	−086	−298
II		−113	149	−002	−253	−357	225	−162	076	108
III			−050	054	−024	042	061	165	112	−008
IV				−139	−302	142	262	−064	−111	061
V					014	137	367	−119	203	−004
VI						068	−128	138	−064	−239
VII							−059	032	017	−116
VIII								−180	102	004
IX									−126	−090
X										001

* Decimals properly preceding each entry have been omitted.

primary trait, namely, Variables 9, 10, 11, and 12 will be discussed in relation to the variables in the present analysis.

Variable 9 in Lurie's analysis is the sum of the scores on the following items: desire to teach modern poetry, to read about art, to read the *Saturday Review,* to have one's son spend his spare time in joining a poetry society, to endow a symphony orchestra, and to listen to a lecture on architecture in relation to the other arts. These various component items might be related to Variables 4, 19, 15, and 27 of the present study although interest in literature as well as interest in other phases of art is indicated in *Variable 9.* This variable would probably be loaded on Factor I if it had been included in the present analysis. The component items in *Variable 10* of Lurie's analysis are as follows: liking for the occupations of musician or portrait painter, tendency to encourage his child to seek a career in literature, and desire to be considered a connoisseur and artistic and talented. No variables in the present analysis resemble any of the items comprising Variable 10. *Variable 11* of Lurie's analysis consists of the following items: interest in Michelangelo, Beethoven, Homer, and Edna St.Vincent Millay; liking for the company of people who are interested in beautiful things; and desire to marry someone who is sensitive and talented. Here again no variables of the present analysis closely resemble any of these component items. *Variable 12* of Lurie's analysis is the sum of the scores of the following items: belief that scientific research should try to make the world more beautiful, that education should teach people to be alive to beauty, that good government should foster the arts, that a good argument against war is that it is dirty and hideous, that a white-collar worker should spend Sunday in attending museums and concerts, that the world needs to replace the ugliness in the world by new beauty. Since Variable 12 contains no components which resemble variables high upon Factor I and since, also, it appears to involve a tendency toward social reform, it is understandable that this variable will not correlate with an aesthetic factor.

In spite of the absence of resemblance between individual items, Lurie's Variables 10 and 11 appear subjectively to measure much the same attributes which are measured by those variables high upon Factor I except, possibly, for the occupational emphasis in Variable 10 which is not apparent in Factor I. If it could be maintained that persons interested in things aesthetic

TABLE 5

LOADING OF THE PRIMARY TRAITS ON THE SECOND-ORDER GENERAL FACTORS (AFTER ROTATION)*

	Primary Trait	*A*	*B*	*C*
I	Aesthetic Evaluative Tendency	366	465	168
II	Interest in Fine Arts	−599	162	−138
IV	Antireligious Evaluative Tendency	−091	−074	−683
V	Antiaggression	060	614	017
VI	Humanitarian Evaluative Tendency	281	077	474
VII	Interest in Science	641	114	−274
VIII	Liberalism	−245	573	−266

* Decimals properly preceding each entry have been omitted.

TABLE 6

The Rotated Factorial Matrix*

Variable	I′	II′	III′	IV′	V′	VI′	VII′	VIII′	IX′	X′	XI′	*A*	*B*	*C*
1	−075	319	112	−050	−105	036	087	099	−018	022	041	−249	033	−085
2	246	451	064	062	−028	−085	009	−054	−119	−129	−146	−261	185	−131
3	126	012	009	074	462	−089	−144	009	−047	−471	166	−084	456	−038
4	333	121	366	024	014	091	−522	049	056	−111	−070	−396	227	294
5	084	096	010	062	190	−209	158	−134	−154	−110	−067	085	129	−209
6	197	−148	−509	401	−051	112	−011	−034	−049	128	149	218	−009	−242
7	281	364	238	−063	093	276	−053	−110	−098	−371	291	−042	287	319
8	286	−107	−137	072	−047	−013	−044	041	326	131	204	166	147	011
9	199	290	078	−234	−023	501	117	−113	067	−089	065	234	182	567
10	−051	−080	066	019	−014	−083	554	018	−038	−056	161	514	033	−311
11	−051	050	043	435	052	098	−094	−066	363	−150	−039	−149	−084	−338
12	−119	143	−129	017	−029	471	029	389	124	−301	044	−119	303	071
13	−232	210	−134	376	273	−096	−063	071	186	−029	−418	−447	126	−548
14	070	022	−024	−365	−048	076	050	409	127	−414	158	−010	407	253
15	214	322	−086	003	073	−020	151	−244	−267	−035	187	072	100	011
16	110	300	044	136	090	−043	403	020	371	055	−103	144	284	−382
17	660	372	−017	066	−094	−324	161	−103	050	044	102	086	348	−225
18	596	−005	049	−210	158	−051	065	−081	−067	−069	−067	433	499	344
19	486	−051	027	076	046	044	−558	076	−216	141	167	−234	323	287
20	070	067	−343	−044	043	−064	050	099	029	001	170	−023	188	−054
21	−017	082	−550	350	037	098	114	−002	399	−119	319	025	030	−361
22	067	425	−051	036	−034	195	038	163	109	074	−023	−275	271	−057
23	267	190	079	162	047	−042	008	010	334	133	081	−047	252	−172
24	022	−007	276	606	223	064	−155	−287	133	124	000	−061	−116	−388
25	089	078	118	040	505	−108	029	199	175	−006	−061	−067	671	−188
26	−021	343	055	578	273	−033	101	009	259	−051	−124	−267	246	−724
27	000	201	−191	−046	112	237	−063	368	032	036	300	−252	469	048
28	−162	080	022	419	078	027	−156	−119	230	062	224	−294	−190	−356
29	420	−187	−070	−083	166	−182	−016	127	277	−055	−020	275	479	066
30	122	−080	−101	291	−098	209	051	−147	−046	−202	049	266	−145	−078
31	197	−041	065	212	118	343	−170	354	012	−006	252	−033	506	001
32	097	−018	102	−217	−099	448	061	329	−065	−159	330	198	329	392
33	052	−051	137	−039	−097	008	287	378	−029	525	265	204	309	−207
34	−010	193	036	−425	−053	065	338	518	076	−054	065	038	531	089
35	502	157	444	377	−080	223	−362	−085	051	113	235	−142	166	041
36	538	071	370	092	−449	−076	−064	−074	−044	426	321	115	−101	022
37	287	−001	−042	062	082	−149	−196	446	059	045	−007	−247	582	−188
38	−002	−189	−035	−065	054	073	122	683	087	005	−158	067	605	−173
39	−017	353	172	−115	144	542	282	086	057	001	130	160	379	246
40	049	008	051	753	034	−021	003	066	043	001	053	−106	038	−806
41	329	106	−069	299	204	104	−202	204	−005	035	252	020	491	−128
42	358	−053	245	334	116	−276	−063	−026	−016	−079	302	046	235	−389
43	−079	441	048	078	294	090	237	011	003	070	202	−142	351	−220
44	209	318	323	−096	186	078	049	246	−064	−065	170	−133	605	029
45	622	246	079	−006	142	−079	−057	064	083	073	165	032	635	063
46	−033	−013	−012	090	374	322	314	119	191	−106	046	387	475	−049
47	199	428	054	−048	−098	−088	−026	158	561	011	172	−370	270	−100
48	−134	109	−072	−053	161	−063	086	−073	642	016	−110	−068	046	−051
49	008	190	137	−029	098	−121	231	433	422	−017	055	−131	525	−357
50	493	115	063	020	−217	062	−123	−094	−281	069	−005	090	077	206
51	−033	−004	−067	−036	347	−022	057	−015	257	039	411	069	277	006
52	−082	−018	092	021	180	096	106	−016	087	061	498	125	109	−007
53	−064	−043	011	776	−065	−034	024	−035	−054	031	024	−084	−221	−827
54	325	217	349	296	061	071	112	344	−123	−182	069	−033	603	−401
55	062	−023	075	−023	489	479	−073	−001	−087	−045	−028	212	500	409
56	−139	−010	007	046	714	060	128	−061	−048	023	−228	152	495	−052
57	419	149	−082	044	050	283	−188	100	224	459	−007	−004	436	260
58	132	−082	452	−239	−023	−299	075	301	103	−249	−127	014	309	−058
59	230	−041	−006	424	123	−133	−015	−099	210	−184	269	083	109	−410
60	−024	005	−265	−031	−170	002	−081	341	014	394	−282	−228	110	−085

* Decimals properly preceding each entry have been omitted.

were no more likely than other individuals to show these occupational preferences, Factor VII of Lurie's analysis,[2] which heavily involves both Variable 9 and Variable 11, would bear some resemblance to Factor I and could plausibly be regarded as an aesthetic factor. If the primary trait involved in these two factors were explained in terms of the traditions of those individuals who have leisure to enjoy the various forms of art, this possibility would appear much more reasonable. Undoubtedly, aesthetic appreciation is associated with the economic standards and leisure of the professional and upper middle classes, whether this is due to the traditions and training of these social classes, to their greater opportunity, or whether it must be explained in some other way.

In summary, then, it may be said that Factor I is undoubtedly aesthetic and that it apparently follows rather closely the content intended by the authors of the test. It will become evident in discussion of Factor II, however, that a considerable proportion of variance of aesthetic variables is not accounted for by Factor I. General Aesthetic Interest will serve as a convenient designation for Factor I.

---

[2] Lurie interprets Factor VII of his analysis as indicating a sort of lip service to aesthetic values. He does not regard Factor VII as well defined or as well established.

Factor II: Interest in the Fine Arts

Variable 2. Loading, .45. He believes that it is justifiable for the great artists such as Beethoven to be selfish and negligent of the feelings of others.

Variable 43. Loading, .44. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician, rather than a sales manager, clergyman, or politician.

Variable 47. Loading, .43. At an evening discussion with intimate friends of his own sex, he is most interested when he talks about philosophy and psychology, literature, socialism and social amelioration rather than the meaning of life.

Variable 22. Loading, .43. He feels that our modern industrial and scientific developments are not signs of a greater degree of civilization and culture than those attained by any previous race, such as the Greeks.

Variable 17. Loading, .37. He would prefer to hear a series of popular lectures on the contemporary painter rather than the progress and needs of social-service work in the cities of his part of the country.

Variable 7. Loading, .36. In a paper such as the *New York Sunday Times,* he is more interested in the section on picture galleries than in the real-estate section and the account of the stock market.

Variable 39. Loading, .35. If he lived in a small town and had more than enough income for his needs, he would prefer to use it to endow the church to which he belonged, give it to a university for the development of scientific research, devote it to hospitals; rather than apply it productively to industrial development.

Variable 26. Loading, .34. Taking the Bible as a whole, he would prefer to regard it from the point of view of its beautiful mythology and literary style rather than as a spiritual revelation.

Variable 15. Loading, .32. If he had the opportunity, and if nothing of the kind existed in the community or college where he lived, he would prefer to found a classical orchestra rather than a debating society.

Variable 1. Loading, .32. He believes that the main object of scientific research should be the discovery of pure truth rather than its practical applications.

Variable 44. Loading, .32. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician, clergyman, or politician; rather than a sales manager.

Variable 16. Loading, .30. At an exposition he prefers to see scientific apparatus or chemical products rather than automobiles.

Factor II appears to be a dimension measuring aspects of aesthetic appreciation and a tendency, often attributed to artists, toward lack of conventional behavior or beliefs. Variables 1, 7, 17, 22, 39, and 43, together possibly with Variables 16 and 44, suggest that the lack of conventionality tends to channelize in negative reactions toward aspects of business and industry. Although some of the negative alternatives are social or religious, none are political.

It should be noted that a liking for reading of *Arts and Decorations* (Variable 19), for collecting paintings and sculpture (Variable 45), belief in the importance of ideals of beauty (Variable 54), liking for ballet (Variable 42), and liking for scenery (Variable 50) are not indicative of high scores on Factor II, although these must doubtless be classified as indicating aesthetic inclinations. Variables 4, 19, and 35 are also uncorrelated with Factor II. Since these are the only items related to literature, the aesthetic tendencies involved in Factor II appear possibly to be limited to the fine arts. The loading of Variable 47 suggests a desire on the part of individuals high on Factor II to appear sophisticated or, at least, to avoid appearing naive, a tendency in keeping with the general tenor of the variables correlating with the factor.

There appears no relation between Factor II and any of the factors isolated by Lurie, nor any consistent relationship to any of the measures obtained by scoring the original test.

The distinction between "we" and "they" groups often employed by sociologists may be important in understanding the determiners of the individual differences measured by Factor II. Persons high on Factor II possibly think of themselves as apart from others expressing conventional attitudes, and their beliefs serve as a means of identification with their group. A large proportion of the responses of an individual high on Factor II might tend to "shock" a conventional person.

Very possibly, the difference between Factors I and II centers in this aspect of Factor II. If, as was suggested in discussing the relation between Factor I and Lurie's Factor II, persons high in socioeconomic status have greater opportunity to develop aesthetic inclination, the emphasis on the "we" and "they" mechanism would appear even more justifiable. An additional and more apparent difference is that Factor II is less general than Factor I and accounts for less variance.

Factor II will tentatively be regarded as measuring "interest in the fine arts."

Factor III: Belief in "Culture"

Variable 21.[3] Loading, .55. He feels that the aim of the churches at the present time should be to convey spiritual worship and a sense of communion with the highest; rather than to bring out charitable tendencies and to urge people to think more of the good of others.

Variable 6.[3] Loading, .51. He considers high ideals and reverence more desirable character traits than unselfishness and sympathy.

Variable 58. Loading, .45. He would prefer, if he were to marry, a wife who could achieve social prestige commanding admiration from others, one who is fundamentally spiritual in her attitude toward life, or one who is gifted along artistic lines; rather than one who likes to stay at home and keep house.

Variable 35. Loading, .44. If he could influence the educational policies of the public schools of some city, he would undertake to promote the study and the performance of the drama; rather than to develop cooperativeness and the spirit of service, to provide additional laboratory facilities, or to promote school savings banks for education in thrift.

Variable 36. Loading, .37. If he could influence the educational policies of the public schools, he would undertake to promote the study and the performance of the drama, to provide additional laboratory facilities, to promote school savings banks for education in thrift; rather than develop cooperativeness and the spirit of service.

Variable 4. Loading, .37. If he were a university professor and had the necessary ability, he would prefer to teach poetry rather than chemistry and physics.

Variable 54. Loading, .35. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's ideals of beauty; rather than one's religious faith, one's business organizations and associates, or society as a whole.

Variable 20.[3] Loading, .34. He would encourage his children to make several teams rather than to have vocational training.

Variable 44. Loading, .32. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician, clergyman, or politician; rather than a sales manager.

Factor III correlates positively with two religious and four aesthetic variables, but does not generally follow the pattern of any of the Allport-Vernon scoring keys.

The common element here involved seems best described as a scale differentiating the person interested in the "higher things of life" from those concerned with a more rational or realistic viewpoint; and persons striving, possibly, for "culture," from those antagonistic toward or suspicious of the intangible nature of such evaluative tendencies.

It should be noted that none of the items which tend to be selected by a person with a positive score on Factor III have obvious, practical, or immediately observable value; none are related to apparent biological or social needs.

---

[3] Both statement and sign of factor loading have been reversed. This reversal allows listing of the loading in positive form.

A person who stresses immediate, tangible benefits can see them resulting from "thinking more of the good of others," "providing laboratories," "promoting thrift or unselfishness," "chemistry," "staying home and keeping house," "vocational training," or "being a sales manager"; but such things as "spiritual worship," "drama," "high ideals and reverence," "poetry," and "ideals of beauty" have value which is difficult to justify on obvious rational grounds. Possibly acceptance of the personal values evident in the variables listed under Factor III depends upon intuitive or mystical processes. Emotional overtones subjectively important to the individual may play an important role. Attention is invited to the considerable proportion of items dealing with ethical problems.

In all, the name "Belief in Culture" characterizes Factor III with reasonable adequacy.

Factor IV: Antireligious Evaluative Tendency

Variable 53. Loading, .78. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's ideals of beauty, one's business organizations and associates, society as a whole; rather than toward one's religious faith.

Variable 40. Loading, .75. If he lived in a small town and had more than enough income for his needs, he would prefer to apply it productively to industrial development, give it to a university for the development of scientific research, or devote it to hospitals; rather than help to endow the church to which he belongs.

Variable 24. Loading, .61. He feels that the world would *not* be a much better place if we took to heart the teaching, "Lay not up for yourselves treasures upon earth . . . but lay up for yourselves treasures in heaven, where neither moth nor rust doth corrupt, and where thieves do not break through nor steal."

Variable 26. Loading, .58. Taking the Bible as a whole he would prefer to regard it from the point of view of its beautiful mythology and literary style rather than as a spiritual revelation.

Variable 11. Loading, .44. When he visits a cathedral, he is more impressed by the architectural features and stained glass than by a pervading sense of reverence and worship.

Variable 34.[4] Loading, .43. He believes that a man who works in business all week can best spend Sunday by trying to win at golf or racing; rather than in trying to educate himself by reading serious books, going to an orchestral concert, or hearing a really good sermon.

Variable 59. Loading, .42. He would like to think of Leonardo da Vinci's "The Last Supper" as one of the most priceless and irreplaceable pictures ever painted, in relation to Leonardo's versatility and its place in history, or as a masterpiece of design; rather than as expressing the highest spiritual tendencies and emotions.

Variable 28. Loading, .42. If he were to choose a topic on which to write an essay, he would choose the defects of our present educational systems rather than the role of church-going in religion.

Variable 6. Loading, .40. He considers unselfishness and sympathy more desirable character traits than high ideals and reverence.

Variable 35. Loading, .38. If he could influence the educational policies of the public schools of some city, he would undertake to promote the study and the performance of drama; rather than to develop cooperativeness and the spirit of service, to provide additional laboratory facilities, or to promote school savings banks for education in thrift.

Variable 13. Loading, .38. He believes that the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first cause, cosmic purpose, or God behind it.

Variable 14.[4] Loading, .37. In his opinion, general progress has been advanced more by the discovery of the steam engine with the consequent industrialization and economic rivalry in American and European countries; rather than by the freeing of slaves with enhancement of the value placed on individual life.

Variable 21. Loading, .35. He feels that the aim of the churches at the present time should be to bring out charitable tendencies, and to urge people to think more of the good of others, rather than to convey spiritual worship and a sense of communion with the highest.

Variable 42. Loading, .33. When he goes to the theatre, he enjoys most, as a rule, ballet or similar imaginative performances; rather than plays that treat the lives of great men, plays with a theme of human suffering and love, or problem plays that argue consistently for some point of view.

With the exception of Variables 3, 9, 18, 38, and 47, all of the religious variables have significant negative loading on Factor IV, and of these five exceptions, Variables 9 and 18 both have loadings greater than .20. While Variables 3, 38, and 47 are regarded as religious by the authors of the test, this interpretation appears questionable. This difference in interpretation, which concerns a possible distinction between particular religious organizations and general religious attitude, will be considered in later discussion of the meaning of this factor.

---

[4] Both statement and sign of factor loading have been reversed.

As in the case of Factor I, the more general interpretation is quite clear. Factor IV is with little doubt antireligious in nature. While its more specific interpretation is not so clear, several possibilities will be discussed in the following paragraphs.

It appears plausible that Factor IV is concerned with the operation of the church as an institution; that it concerns the creed and social controls of the Christian church; and that it does not measure any fundamental evaluative trait common to religious individuals of other than the Christian creed. (It will be convenient to speak of religious evaluation even though *low* scores are indicative of *high* religious evaluative tendency.) In the absence of any direct indication to the contrary, "Christian" will refer to both the Protestant and Catholic churches. Those items loaded to the extent of .45 or greater most certainly give the impression of measuring mere conventional indoctrination. In numerous instances the social forces which operate to ostracize the doubter and disbeliever appear to be determining the responses. The aforementioned insignificant loadings of Variables 3, 38, and 47, which were apparently included by the authors of the test in an attempt to broaden the measure of religious evaluation beyond that of the conventional church member, are important to this interpretation.

Further support of this theory is found in the loading of certain variables on the religious factor appearing in Lurie's analysis, in that variables with component items designed to measure a religious evaluative tendency more general than that involved in the Christian religion did not have high loadings on his religious factor. It is possible that if these component items measuring religious evaluative tendency non-Christian in nature could have been separated from the remaining items or those measuring Christian religious evaluation, that such non-Christian items would have been found to correlate negatively with his religious factor. Variable 22, wherein the items all concern conventions of the Christian religion, is the only variable in Lurie's analysis heavily involving his religious factor.

Certain further elaborations upon the simple hypothesis of religious tendency are suggested. Since agreement with such phases as "the aim of the churches at the present time should be to bring out charitable tendencies, and to urge people to think more of the good of others" (Variable No. 21), schools should develop "cooperativeness and the spirit of service" (Variable No. 36), or "he considers unselfishness and sympathy more desirable character traits than high ideals and reverence" (Variable No. 6) is always indicative of nonreligious tendency insofar as this may be determined by Factor IV, it is not unreasonable to conclude that religious tendency as defined by Factor IV is not primarily ethical in nature. Of course, two of these above-listed statements, namely those from Variable 6 and 21, involve a choice between ethical values which should be related to religious evaluative tendency. However, in the remaining instance, that from Variable 36, even though no choice was involved between ethical and other aspects of religious evaluation, the same negative correlation was observed.

A consideration of Variables 6, 11, 21, 26, and 59 leads to the conclusion that reverence is an important aspect of religious evaluation as defined by Factor IV. This is in agreement with the previously suggested interpretation of Factor IV as a measure of the tendency to accept the teachings of the church without question.

In summary, then, Factor IV is undoubtedly antireligious in nature and is probably concerned largely with evaluation of the church as an institution. Factor IV will be referred to as "Antireligious Evaluative Tendency."

Factor V: Antiaggression

Variable 56. Loading, .71. Florence Nightingale, Henry Ford, or Charles Darwin interests and attracts him more than Napoleon.

Variable 25. Loading, .51. He is more interested in reading accounts of the lives and works of such men as Aristotle, Plato, and Socrates than of Alexander, Julius Caesar, and Charlemagne.

Variable 55. Loading, .49. Florence Nightingale interests and attracts him more than Napoleon, Henry Ford, or Charles Darwin.

Variable 3. Loading, .46. He does not believe that because of the aggressive and self-assertive nature of man the abolition of war is an illusory ideal.

Variable 36.[5] Loading, .45. If he could influence the educational policies of the public schools, he would undertake to develop cooperativeness and the spirit of service; rather than to promote the study and the performance of the drama, to provide additional laboratory facilities, or to promote school savings banks for education in thrift.

---

[5] Both statement and sign of factor loading have been reversed.

Variable 46. Loading, .37. If he had unlimited leisure and money, he would prefer to establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient; rather than make a collection of fine sculptures or paintings, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

Variable 51. Loading, .35. Great exploits and adventures of discovery such as Lindbergh's and Byrd's seem significant because they add to our knowledge of mechanics, geography, meteorology, etc., weld human interests and international feelings throughout the world, or contribute to the ultimate revelation of the meaning of the universe; rather than because they represent conquests by man over the difficult forces of nature.

Variable 43. Loading, .29. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician; rather than a sales manager, clergyman, or politician.

While the relation of the Allport-Vernon theoretic score to Factor V is not so clearly evident as in the case of Factors I and IV, there are certain resemblances which will bear comment. Of the 8 variables with significant loading on Factor V, 4 were political, and of the remaining 3, 2 involved political evaluations as alternate choices. All of such political evaluations correlated negatively with Factor V. On the positive pole, 3 of these above-listed variables were social and 2 alternative choices were social. One of the variables is theoretical and 4 additional variables involved alternative choices similarly designated. It appears that Factor V may be best described in terms of the evaluative scores as social, and possibly theoretic on the positive pole and as political on the negative pole.

Examination of the listing of variables with high loading on Factor V suggests that, on the negative pole, the interests and evaluations expressed seem related to conflict, possibly even more to armed conflict, while on the positive side, the intellectual interests and social sympathies which seem to be expressed are somewhat withdrawing in nature. It should be noted that none of the theoretic items involved in Factor V are measures of interest in the natural sciences. None of the items high on Factor V scored as social appear to concern social issues which are in active dispute at the present time.

The term "Antiaggression" will serve to identify Factor V.

Factor VI: Humanitarian Tendency

Variable 39. Loading, .54. If he lived in a small town and had more than enough income for his needs, he would prefer to use it to endow the church to which he belonged, give it to a university for the development of scientific research, devote it to hospitals; rather than apply it productively to industrial development.

Variable 9. Loading, .50. He should prefer to read the headlines "Protestant Leaders To Consult On Reconciliation" rather than "Great Improvements In Market Conditions."

Variable 55. Loading, .48. Florence Nightingale interests and attracts him more than Napoleon, Henry Ford, or Charles Darwin.

Variable 12. Loading, .47. He does not believe that contemporary charitable policies should be curtailed because they tend to undermine individual initiative.

Variable 32. Loading, .45. He believes that good government should aim at more aid for the poor, sick, and old, the introduction of more ethical principles into its policies and diplomacy, the establishment of a position of prestige and respect among nations; rather than the development of manufacturing and trade.

Variable 31. Loading, .34. He believes that good government should aim at more aid for the poor, sick and old; rather than the development of manufacturing and trade, introduction of more ethical principles into its policies and diplomacy, or the establishing of a position of prestige and respect among nations.

Variable 17.[6] Loading, .32. He would prefer to hear a series of popular lectures on the progress and needs of social-service work in the cities of his part of the country rather than on the contemporary painter.

---

[6] Both statement and sign of factor loading have been reversed.

Variable 46. Loading, .32. If he had unlimited leisure and money, he would prefer to establish a mental hygience clinic for taking care of the maladjusted and mentally deficient; rather than make a collection of fine sculptures or paintings, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

On the positive pole, Factor VI heavily involves social variables and correlates negatively with a number of items measuring economic evaluative tendency. An appreciable number of items measuring theoretic evaluative tendency correlate with Factor VI. While there might seem to be some resemblance to Factor V because of the large number of social items involved in each instance, they differ from each other in that social items of Factor V are largely concerned with evaluations of persons or of characteristics of persons while those high on Factor VI tend more to be concerned with social issues. Possibly Factor VI can best be described as concerned with interest in social reform that will aid in improving the welfare of the underprivileged. The term Humanitarian Tendency appears most appropriate of those occurring to the author.

In the above listing only Variable 9 appears at all inconsistent with this statement as to the nature of Factor VI. Even here the inconsistency is not so much in its loading on Factor VI but in the size of the loading. With the possible exception of Variables 6 and 21, none of the variables having near zero loading on Factor VI appear to measure the primary trait described in the above hypothesis. The abstract statement of Variables 6 and 21 and the fact that both involve a religious evaluation as an alternative to an apparent measure of a humanitarian tendency, may help to explain their lack of loading on Factor VI.

Factor VI might be regarded, on the surface, as resembling Lurie's Factor I, which he interpreted as being social in nature. However, in spite of the many items high on Factor VI which were intended as a measure of the "social man," a close comparison of the individual items of Factor VI reveals little definite resemblance.

Variables 14 and 27 are the only apparent measures of humanitarianism not loaded on Factor VI. Of these two, Variable 27 has a loading of .237.

The term "Humanitarian Tendency" will be used to indentify Factor VI.

Factor VII: Interest in Science

Variable 19.[7] Loading, .56. If he had some time to spend in a waiting room and there were only two magazines to choose from, he would read *The Scientific American* rather than *Arts and Decorations.*

Variable 10. Loading, .55. He would prefer to read the story "Physicists Make Important Discoveries Concerning The Structure Of Atoms" rather than "Threat To Constitutional Government In Germany By A Dictator."

Variable 4.[7] Loading, .52. If he were a university professor and had the necessary ability, he would prefer to teach chemistry and physics rather than poetry.

Variable 16. Loading, .40. At an exposition he prefers to see scientific apparatus or chemical products rather than automobiles.

Variable 35.[7] Loading, .36. If he could influence the educational policies of the public schools of some city, he would undertake to develop cooperativeness and the spirit of service, to provide additional laboratory facilities, or to promote school savings banks for education in thrift; rather than undertake to promote the study and performance of drama.

Variable 34. Loading, .34. He believes that a man who works in business all week can best spend Sunday in trying to educate himself by reading serious books, going to an orchestral concert, or hearing a really good sermon; rather than by trying to win at golf or racing.

Variable 46. Loading, .31. If he had unlimited leisure and money, he would prefer to establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient; rather than make a collection of fine sculptures or paintings, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

All except three of the variables significantly loaded on Factor VII clearly imply an interest in the sciences. Furthermore, with the possible exception of Variable 49, which has a loading of .231, there are no additional items in the present analysis not loaded on Factor VII which can readily be interpreted as measuring such interest. Hence, there appears little doubt that Factor VII is a primary trait measuring interest in science.

---

[7] Both statement and sign of factor loading have been reversed.

The loadings of the three items not obviously concerned with interest in science, namely, numbers 34, 35, and 46, are not necessarily inconsistent with this hypothesis as to the nature of Factor VII. It should be noted that one of the alternatives of Variable 35 is concerned with the promotion of laboratory work in the schools. This may partly explain its loading on Factor VII. Possibly, with reference to Variable 34, it is characteristic of scientists that they dislike the idea of promoting the study of drama in the public schools.

The isolation of Factor VII, which involves only a few of the theoretical items of the present analysis, clearly indicates the existence of several rather than a single primary trait of theoretical evaluative tendencies. Lurie's findings of a single primary trait of theoretical interest is probably due to the complexity of the variables in his analysis. (It will be remembered that each of his variables was the sum of several items.) It is interesting in this connection that a minus rotation, in Lurie's analysis, of Factor V on Factor I would leave only Variables 1 and 3 of his analysis high on Factor V. His Factor V would, then, bear some resemblance to Factor VII of the present analysis.

Factor VII will be identified by the term, "Interest in Science."

Factor VIII: Tendency toward Liberalism

Variable 38. Loading, .68. He would prefer a friend of his own sex who is seriously interested in thinking out his attitude toward life as a whole; rather than one who is efficient, industrious, and of a practical turn of mind, one who possesses qualities of leadership and organizing ability, or one who shows refinement and emotional sensitivity.

Variable 34. Loading, .52. He believes that a man who works in business all week can best spend Sunday in trying to educate himself by reading serious books, going to an orchestral concert, or hearing a really good sermon; rather than by trying to win at golf or racing.

Variable 37. Loading, .45. He would prefer a friend of his own sex who is seriously interested in thinking out his attitude toward life as a whole, one who possesses qualities of leadership and organizing ability, one who shows refinement and emotional sensitivity; rather than one who is efficient, industrious, and of a practical turn of mind.

Variable 49. Loading, .43. During his next summer vacation he would prefer to write and publish an original biological essay or article; rather than stay in some secluded part of the country where he could appreciate fine scenery, go in for a local tennis or other athletic tournament, or get experience in some new line of business.

Variable 14. Loading, .41. In his opinion, general progress has been advanced more by the freeing of slaves with enhancement of the value placed on individual life rather than by the discovery of the steam engine with consequent industrialization and economic rivalry in American and European countries.

Variable 12. Loading, .39. He does not believe that contemporary charitable policies should be curtailed because they tend to undermine individual initiative.

Variable 33. Loading, .38. He believes that a man who works in business for his living all the week can best spend Sunday in trying to educate himself by reading serious books; rather than by trying to win at golf or racing, going to an orchestral concert, or by hearing a really good sermon.

Variable 27. Loading, .37. He does not believe that since the class or social status to which a man belongs depends mainly upon his push and ability, it is just that a small proportion of the population should be very rich.

Variable 31. Loading, .35. He believes that good government should aim at more aid for the poor, sick, and old; rather than the development of manufacturing and trade, introduction of more ethical principles into its policies and diplomacy, or the establishing of a position of prestige and respect among nations.

Variable 54. Loading, .34. He believes that one should guide one's conduct according to, or develop one's chief loyalties towards, one's ideals of beauty; rather than one's religious faith, one's business organizations and associates, or society as a whole.

Variable 60. Loading, .34. He would like to think of Leonardo da Vinci's picture "The Last Supper" as expressing the highest spiritual tendencies and emotions, in relation to Leonardo's versatility and its place in history, as a masterpiece of design; rather than as one of the most priceless and irreplaceable pictures ever painted.

Variable 32. Loading, .33. He believes that good government should aim at more aid for the poor, sick, and old, the introduction of more ethical principles into its policies and diplomacy, the establishment of a position of prestige and respect among nations; rather than the development of manufacturing and trade.

Variable 58. Loading, .30. He would prefer, if he were to marry, a wife who could achieve social prestige, commanding admiration from others, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who likes to stay at home and keep house.

Factor VIII heavily involves items intended as measures of the economic man on its negative pole and correlates positively with items intended as measures of the social and of the theoretic man. Thus, there are 7 economic variables high on Factor VIII and 4 with economic alternatives. On the other hand, the principal items in 4 of the above-listed variables and alternatives in 2 additional variables are social measures. Two variables and alternatives of 3 additional variables loaded upon Factor VIII are theoretic measures. Variables 5, 7, 9, 20, 22, 24, 29, 39, and 44 comprise a listing of those economic variables not significantly loaded on Factor VIII. If we neglect Variables 5 and 20 on the ground that they are not loaded significantly on any of the primary traits of the present analysis, and Variables 22 and 24 on the ground that their loadings approach significance, and compare the remaining above-listed variables, namely, numbers 7, 9, 29, and 39 with numbers 12, 14, 37, 32, 38, 27, and 28 of the list of variables high on Factor VIII, it seems apparent that the economic items loaded on Factor VIII may be differentiated from those not so loaded in that they either state or imply the desirability or undesirability of characteristics of our economic system. Further evidence of a similar nature is found in examining the afore-mentioned social and economic variables significantly loaded on Factor VII. These may likewise be differentiated from social and theoretic variables not loaded on Factor VIII in that the former are concerned with social issues whereas the latter are not. By and large these social and theoretic items might be described as expressing beliefs favoring the underprivileged groups or possibly as being "liberal" or "New Dealish." This may not be so apparent in the case of Variables 34, 37, 38, and 49. However, these can be explained by the plausible assumption that beliefs concerning economic systems are in part, at least, a function of social groupings and the ways of life of these social groups.

If we assume similar tendencies on the part of the "liberals" at the opposite pole, we may readily understand the loadings of Variables 38 and 49. Since Variables 34, 37, 38, and 49 all have higher-than-average loadings, it would seem that such preferences are themselves better-than-average measures of liberal tendencies and are more fundamental variables in the cluster of traits defined by this factor.

The term "Tendency toward Liberalism" will be employed to designate Factor VIII. It is suggested that the factor measures differences between social groups such as business men on the one hand and "liberals" or "New Dealers" on the other.

While Lurie's Factor II (which he labeled a measure of philistine tendencies) bears rather general resemblance to Factor VIII, more specific examination of the variables defining the two factors suggests that the resemblance is superficial. Thus, Lurie's Factor II is correlated with variables measuring interest in politics and with others that he had regarded as aesthetic. These latter are similar to those highly loaded on

Factor II of the present analysis. A further difference is found in the lack of social or theoretic variables in Lurie's Factor II, although the resemblance of Lurie's "social" and "theoretic" variables to those of the present analysis is so uncertain as to make comment on this discrepancy rather difficult.

Since Factor VI appeared also to involve a liberal or progressive point of view, certain differences between these factors will be indicated. The high loading of Variables 17, 46, and 55 on Factor VI and their near zero loadings on Factor VIII suggest that the personal and emotional component of Factor VI is absent in Factor VIII. By contrast, Factor VIII is impersonal and intellectual. Similarly, the loadings of Variables 9, 39, and possibly 7 suggest that the person high on Factor VI is likely to show no interest in business or industry while the individual high on Factor VIII is apparently as likely as the person low on Factor VIII to show interest in business trends. The distinguishing characteristic of the person high on Factor VIII is his disagreement with the evaluations of business men, but not in his lack of interest in economic problems.

Factor IX: Theoretic Interest

Variable 48. Loading, .64. At an evening discussion with intimate friends of his own sex, he is most interested when they talk about philosophy and psychology; rather than the meaning of life, literature, or socialism and social amelioration.

Variable 47. Loading, .56. At an evening discussion with intimate friends of his own sex, he is most interested when he talks about philosophy and psychology, literature, socialism and social amelioration; rather than the meaning of life.

Variable 49. Loading, .42. During his next summer vacation, he would prefer to write and publish an original biological essay or article; rather than stay in some secluded part of the country where he could appreciate fine scenery, go in for a local tennis or other athletic tournament, or get experience in some new line of business.

Variable 21. Loading, .40. He feels that the aim of the churches at the present time should bring out charitable tendencies, and urge people to think more of the good of others; rather than to convey spiritual worship and a sense of communion with the highest.

Variable 16. Loading, .37. At an exposition he prefers to see scientific apparatus or chemical products rather than automobiles.

Variable 11. Loading, .36. When he visits a cathedral, he is more impressed by the architectural features and stained glass than by the pervading sense of reverence and worship.

Variable 23. Loading, .33. If he were waiting in a strange living room, he would be more apt to judge the owner's knowledge and intellectuality as shown by the books in his bookcases than his friendliness and hospitality as shown by the photographs on the walls and the general comforts of the room.

Variable 8. Loading, .33. He does not think that a person who analyzes his emotions is apt to be less sincere in his feeling than one who is not so reflective.

Of the above listing, Variables 48, 49, 16, 23, and 8 are all scored as theoretic. If we exclude from consideration the several items involved in Factor VII, the interest in science factor: viz., Variables 15, 52, and possibly 25, in that these have marginal loadings on Factor IX; and Variables 15, 52, and possibly 25, in that they have marginal loadings on Factor IX, then Variables 1, 33, and 43 are the only remaining measures of theoretic evaluative tendency not loaded on Factor IX. Although Variables 11 and 21, both of which are significantly loaded on Factor IX, contain no items intended as measures of theoretic tendency, it is the author's opinion that these variables are nonetheless indicative of theoretic interests. The alternative items in Variable 47 undoubtedly explain the findings that this variable correlates with an apparent measure of theoretic tendency. Consequently, Factor IX will be regarded as a measure of theoretic interests. The theoretic element common to the variables having high loadings on Factor IX is identified with somewhat more confidence than in the case of a number of the remaining factors, since the variables concern a considerable variety of activities and beliefs.

It has already been mentioned that Lurie isolated a single "theoretic" primary evaluative tendency. The relation of this trait to Factors VII and IX of the present analysis is not clear. Probably the present analysis, with less complex variables, has succeeded in breaking Lurie's single primary trait into two separate factors. Neither of these two factors bears striking resemblance to Lurie's theoretical factor. In part, this is doubtless due to the earlier mentioned fact that Lurie's theoretic (and social) variables are difficult to relate to those of the Allport-Vernon test.

Factor X: "Rugged Individualism"

Variable 33. Loading, .53. He believes that a man who works in business for his living all the week can best spend Sunday in trying to educate himself by reading serious books; rather than by trying to win at golf, or racing, going to an orchestral concert, or hearing a really good sermon.

Variable 3.[8] Loading, .47. He believes that because of the aggressive and self-assertive nature of man the abolition of war is an illusory ideal.

Variable 57. Loading, .46. He would prefer, if he were to marry, a wife who would like to stay at home and keep house, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who can achieve social prestige commanding admiration from others.

Variable 36. Loading, .43. If he could influence the educational policies of the public schools, he would undertake to promote the study and the performance of the drama, to provide additional laboratory facilities, to promote school savings banks for education in thrift; rather than to develop cooperativeness and the spirit of service.

Variable 14.[8] Loading, .41. In his opinion, general progress has been advanced more by the discovery of the steam engine with the consequent industrialization and economic rivalry in American and European countries, than by the freeing of slaves with enhancement of the value placed on individual life.

Variable 60. Loading, .39. He would like to think of Leonardo da Vinci's picture "The Last Supper" as expressing the highest spiritual tendencies and emotions, in relation to Leonardo's versatility and its place in history, as a masterpiece of design; rather than as one of the most priceless and irreplaceable pictures ever painted.

Variable 7.[8] Loading, .37. In a paper such as the *New York Sunday Times,* he is more interested in the real-estate section and the account of the stock market than in the section on picture galleries.

Variable 12.[8] Loading, .30. He believes that contemporary charitable policies should be curtailed because they tend to undermine individual initiative.

When the rotations from the centroid to the primary-trait loadings were first thought to be complete, Factor X appeared of doubtful significance. Apart from the afore-mentioned difficulties in the rotation of Factor X, its high saturation with the later centroid factors should suggest that its significance as a primary trait be accepted rather tentatively. Consequently, further rotations were made in an attempt to produce a more meaningful configuration. Although those items with loadings of borderline significance could be altered, those with loading above .4 maintained their identity throughout these additional rotations. More careful examination suggested that the common elements to the variables of the above listing could be meaningfully interpreted.

Factor X appears to be unrelated to any of the six original scores of the test or to any of the factors isolated by Lurie.

Tentatively, it is suggested that Factor X is related to the American tradition of rugged individualism; to the belief that people get ahead through hard work and are motivated by competition and desire for economic advancement. It is emphasized that this tradition as here interpreted has been closely associated with and advocated by business and industry. This hypothesis directly explains the loadings of Variables 33, 7, 12, and 14. The belief in the inevitability of war expressed in Variable 3 is probably a generalization of the idea that man is competitive by nature. Variables 5 and 7 may reflect the traditional middle-class male dislike of society, a dislike which would seem related to rugged individualism, while the loading of Variable 60 may express a revolt of such an individual against what he considers an "over-evaluation" of da Vinci's painting. It should be noted that the three alternative choices of Variable 60 suggest somewhat less enthusiastic evaluation.

---

[8] Both statement and sign of factor loading have been reversed.

Factor X will be considered a measure of adherence to the evaluative standards associated with the term "Rugged Individualism."

Factor XI: Residual in Nature

Variable 52. Loading, .50. Great exploits and adventures of discovery such as Lindbergh's and Byrd's seem significant to him because they add to our knowledge of mechanics, geography, meteorology, etc.; rather than because they represent conquests by man over the difficult forces of nature, because they weld human interests and international feelings throughout the world, or because they contribute to the ultimate revelation of the meaning of the universe.

Variable 13.[9] Loading, .42. He does not believe that the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first cause, cosmic purpose, or God behind it.

Variable 51. Loading, .41. Great exploits and adventures of discovery such as Lindbergh's and Byrd's seem significant because they add to our knowledge of mechanics, geography, meterology, etc., because they weld human interests and international feelings throughout the world or because they contribute to the ultimate revelation of the meaning of the universe; rather than because they represent conquests by man over the difficult forces of nature.

Variable 32. Loading, .33. He believes that good government should aim at more aid for the poor, sick, and old, the introduction of more ethical principles into its policies and diplomacy, the establishment of a position of prestige and respect among nations; rather than the development of manufacturing and trade.

Variable 36. Loading, .32. If he could influence the educational policies of the public schools, he would undertake to promote the study and the performance of the drama, to provide additional laboratory facilities, to promote school savings banks for education in thrift; rather than to develop cooperativeness and the spirit of service.

Variable 21. Loading, .32. He feels that the aim of the churches at the present time should be to bring out charitable tendencies, and to urge people to think more of the good of others; rather than to convey spiritual worship and a sense of communion with the highest.

Variable 42. Loading, .30. When he goes to the theater, he enjoys most, as a rule, ballet or similar imaginative performances; rather than plays that treat the lives of great men, plays with a theme of human suffering and love, or problem plays that argue consistently for some point of view.

Variable 27. Loading, .30. He does not believe that since the class or social status to which a man belongs depends mainly upon his push and ability, it is just that a small proportion of the population should be very rich.

No very consistent tendency is apparent in the meaning of variables with high loading on Factor XI. Since, in addition, few items have significant loadings, and the total variance accounted for is small, and the projections of this factor are heaviest on the later centroid factors (which are heavily saturated with sampling error), it seems reasonable to conclude that this factor is residual in nature. While it is often considered advisable to rotate the reference axis of the residual factor so that it is orthogonal to the remaining reference axes, it is evident from Table 4 that a minus rotation on Factor I, which will for practical purposes achieve orthogonality between Factor XI and the remaining primary traits, will produce no substantial changes in the variables high on Factor XI.

C. The Second-Order Factor

It will be recalled from discussions

---

[9] Both statement and sign of factor loading have been reversed.

under "Procedure" that three second-order factors were extracted from the intercorrelations of the primary traits and that the loadings of the test items on these second-order factors were then computed. The interpretation of these second-order factors will now be considered. In Table 6 these factors were labeled *A, B,* and *C* to distinguish them from the eleven first-order factors.

Second-Order Factor A*: Not Identified*

Variable 10. Loading, .51. He would prefer to read the story "Physicists Make Important Discoveries Concerning The Structure of Atoms" rather than "Threat To Constitutional Government In Germany By A Dictator."

Variable 13.[10] Loading, .45. He does *not* believe that the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first cause, cosmic purpose, or God behind it.

Variable 18. Loading, .43. He would prefer to hear a series of lectures on the comparative development of the great religious faiths rather than the comparative merits of the forms of government in Britain and in the United States.

Variable 4.[10] Loading, .40. If he were a university professor and had the necessary ability, he would prefer to teach chemistry and physics rather than poetry.

Variable 46. Loading, .39. If he had unlimited leisure and money, he would prefer to establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient; rather than make a collection of fine sculptures or paintings, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

Variable 47.[10] Loading, .37. At an evening discussion with intimate friends of his own sex, he is more interested when he talks about the meaning of life; than when he talks about philosophy and psychology, literature, or socialism and social amelioration.

Factor *A* is difficult to interpret. Since many of the first-order primary traits were not as clearly defined as might be desired, the intercorrelations of these traits are probably more unstable than would be expected from sampling error alone. For this reason it would seem inadvisable to concern ourselves very greatly with the meaning of factors such as *A* which do not, in any case, account for a large proportion of the variance.

Second-Order Factor B*: Idealism vs. Practicality*

Variable 25. Loading, .67. He is more interested in reading accounts of the lives and works of such men as Aristotle, Plato, and Socrates; than of such men as Alexander, Julius Caesar, and Charlemagne.

Variable 45. Loading, .64. If he had unlimited leisure and money, he would prefer to make a collection of fine sculptures or paintings; rather than establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

Variable 44. Loading, .61. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician, clergyman or politician; rather than a sales manager.

Variable 38. Loading, .61. He would prefer a friend of his own sex who is seriously interested in thinking out his attitude toward life as a whole; rather than one who is efficient, industrious, and of a practical turn of mind, one who possesses qualities of leadership and organizing ability, or one who shows refinement and emotional sensitivity.

Variable 54. Loading, .60. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's ideals of beauty; rather than one's religious faith, one's business organizations and associates, or society as a whole.

Variable 37. Loading, .58. He would prefer a friend of his own sex who is seriously interested in thinking out his attitude toward life as a whole, one who possesses qualities of leadership and organizing ability, one who shows refinement and emotional sensitivity; rather than one who is efficient, industrious, and of a practical turn of mind.

Variable 34. Loading, .53. He believes that a man who works in business all week can best spend Sunday in trying to educate himself by reading serious books, going to an orchestral concert, or hearing a really good sermon; rather than by trying to win at golf or racing.

Variable 49. Loading, .53. During his next summer vacation he would prefer to write and publish an original biological essay or article;

---

[10] Both statement and sign of factor loading have been reversed.

rather than stay in some secluded part of the country where he could appreciate fine scenery, compete in a local tennis or other athletic tournament, or get experience in some new line of business.

Variable 31. Loading, .51. He believes that good government should aim at more aid for the poor, sick, and old; rather than the development of manufacturing and trade, introduction of more ethical principles into its policies and diplomacy, or the establishing of a position of prestige and respect among nations.

Variable 55. Loading, .50. Florence Nightingale interests and attracts him more than Napoleon, Henry Ford, or Charles Darwin.

Variable 18. Loading, .50. He would prefer to hear a series of lectures on the comparative development of the great religious faiths rather than the comparative merits of the forms of government in Britain and the United States.

Variable 56. Loading, .50. Florence Nightingale, Henry Ford, or Charles Darwin interests and attracts him more than Napoleon.

Variable 41. Loading, .49. When he goes to the theatre he enjoys most, as a rule, the ballet or similar imaginative performances, plays with a theme of human suffering and love, problem plays that argue consistently for some point of view; rather than plays that treat the lives of great men.

Variable 29. Loading, .48. If he were to choose a topic on which to write an essay, he would prefer the personality of some close friend of his rather than the necessities of life, luxuries, and savings.

Variable 46. Loading, .48. If he had unlimited leisure and money, he would prefer to establish a mental hygiene clinic for taking care of the maladjusted and mentally deficient; rather than make a collection of fine sculptures or paintings, aim at a senatorship or a seat in the Cabinet, or enter into banking and high finance.

Variable 27. Loading, .47. He does not believe that, since the class or social status to which a man belongs depends mainly upon his push and ability, it is just that a small proportion of the population should be very rich.

Variable 3. Loading, .46. He does not believe that because of the aggressive and self-assertive nature of man the abolition of war is an illusory ideal.

Variable 57. Loading, .44. He would prefer, if he were to marry, a wife who likes to stay at home and keep house, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who can achieve social prestige commanding admiration from others.

Variable 14. Loading, .41. In his opinion, general progress has been advanced more by the freeing of slaves with enhancement of the value placed on individual life rather than by the discovery of the steam engine with the consequent industrialization and economic rivalry in American and European countries.

Variable 39. Loading, .38. If he lived in a small town and had more than enough income for his needs, he would prefer to help to endow the church to which he belonged, give it to a university for the development of scientific research, devote it to hospitals; rather than apply it productively to industrial development.

Variable 43. Loading, .35. Assuming that he had the necessary ability and that the salary for each of the following occupations was the same, he would prefer to be a mathematician; rather than a sales manager, clergyman, or politician.

Variable 17. Loading, .35. He would prefer to hear a series of popular lectures on the contemporary painter rather than on the progress and needs of social service work in the cities of his part of the country.

Variable 32. Loading, .33. He believes that good government should aim at more aid for the poor, sick, and old, the introduction of more ethical principles into its policies and diplomacy, the establishment of a position of prestige and respect among nations; rather than the development of manufacturing and trade.

Variable 19. Loading, .32. If he had some time to spend in a waiting room and there were only two magazines to choose from, he would read *Arts and Decorations* rather than *The Scientific American.*

Variable 58. Loading, .31. He would prefer, if he were to marry, a wife who could achieve social prestige commanding admiration from others, one who is fundamentally spiritual in her attitude toward life, one who is gifted along artistic lines; rather than one who likes to stay home and keep house.

Variable 33. Loading, .31. He believes that a man who works in business for his living all week can best spend Sunday in trying to educate himself by reading serious books; rather than by trying to win at golf or racing, going to an orchestral concert, or hearing a really good sermon.

Variable 12. Loading, .30. He does not believe that contemporary charitable policies should be curtailed because they tend to undermine individual initiative.

In terms of the content of the variables as shown by the Allport-Vernon scoring key for the Test of Values, Factor *B* is heavily antipolitical and antieconomic. To a somewhat lesser degree content scored as social, aesthetic, and theoretic is evident among the alternatives having positive correlation with Factor *B*. A number of the variables involving alternatives scored as religious have high loading on Factor *B*, notably variables 38, 18, and 3. However, these would not, in the author's opinion, usually be considered indicative of religious evaluative tendency. It is pertinent to note that none of these variables had significant projections on first-order Factor III, which had been interpreted as primarily antireligious in content.

Factor *B* resembles Lurie's Philistine factor (5). Detailed examination of the variables with high loadings on Factor *B* reveals no marked inconsistencies between their content and the term Philistine. However, the author feels that other designations are more appropriate and prefers to avoid the emotional tone of the term "Philistine."

Persons scoring high on Factor *B* may be characterized as idealistic, if the following restricted interpretation of the term is kept in mind. Idealistic, as we will use the term, characterizes individuals who evaluate highly concepts and beliefs which are aimed at the solution of the problem of the ideal social structure, the ideal manner of life, and who evaluate highly those aspects of our present day and of past cultures which they believe to be improvements over the usual, accepted, or conventional set of values. In line with this tendency, the individual high on Factor *B* appears to evaluate contemplation more highly than direct action. This latter tendency is further evidenced in aesthetic preferences that the more practical individual would regard as esoteric, involved, or "high brow."

It was indicated in foregoing discussion that items on the positive pole were, according to the scoring key for the test, aesthetic, social, and theoretic in content. The comments in the preceding paragraph suggest a reasonable hypothesis as to the common component in high evaluation of these apparently different types of content. The first two of these three types of content suggest idealism with respect to cultural development of the individual, and with respect to human relations and the structure of society. Idealism, as we use the term, is logically related to high evaluation of theoretic items, since speculation, contemplation, or theorizing is a necessary activity of individuals who concern themselves with "how things should be" rather than accept "things as they are." There is the further suggestion that individuals scoring high on Factor *B* may evaluate more highly what they have acquired through the schools, through reading, and through contact with what might be termed cultural influences. They appear, in other words, to have been more heavily influenced by such aspects of our present-day culture than by the everyday world of practical affairs.

Individuals who are low scoring or on the negative pole of Factor *B* are characterized in part by a tendency toward low evaluation of variables with social, aesthetic, and theoretic content. Such individuals appear to be unconcerned with the problem of "how things should be" but accept the "things as they are" and evaluate favorably those things which are related to effective action within the framework of the accepted set of values of the dominant "we group" of our present-day culture. On a somewhat more tenuous basis, the author suggests that the low-scoring individuals, because they accept "things as they are" and are concerned with effective action within this framework, react with antagonism, lack of patience, lack of understanding, or disparagement toward refusal to accept "things as they are."

The foregoing comments suggest the desirability of a bipolar designation for Factor *B*. Since a factor is a single dimension, the need for a bipolar type of description requires justification. In normal interpretation of factors it is assumed that individuals scoring low lack characteristics possessed to a greater degree by individuals scoring higher. With a bipolar description it is implied that individuals high on the factor are characterized by evaluative tendencies absent in those scoring low and conversely that those scoring low are characterized by evaluative tendencies absent in those scoring high on the factor. The author believes that the bipolar nature of Factor *B* results from the interaction of social groups with which the "Philistine" and "idealist" identify themselves—mutual disparagement tends to develop. Individuals possessing the characteristics of the high- or low-scoring group tend in consequence not to possess the evaluative tendencies characterizing those at the opposite pole.

As an additional point we might note that the content of the variables with high loading on Factor *B* suggests the presence of withdrawal tendencies on the part of both high- and low-scoring individuals. The high-scoring individuals appear to show tendencies toward withdrawal from active participation in the usual or dominant social structure. It appears possible, actually, that Factor *B* may show correlation with such personality factors as Moiser's (4) and Guilford's (3) social-introversion factor. On the other hand, individuals scoring low on Factor *B* appear to show withdrawal from "cultural" activities. It is difficult in both instances to decide whether the apparent withdrawal is evidence of preference for the more highly evaluated activities or whether actual withdrawal is indicated.

Whatever its exact nature, the author believes that Factor *B* is an important source of variance in the value judgments of human beings and that its significance carries over into areas beyond that of value judgments. Evidence in support of this statement may be found in the table of intercorrelation of occupational interest scores published by E. K. Strong (6). In this intercorrelation table, a cluster of occupational scores fitting the general definitions of the two poles of Factor *B* can be readily selected. Equivalents of the positive pole may be found in scores for mathematician, artist, and the sciences. On the negative pole such occupations as sales manager and banker are to be found. The average intercorrelation within the positive cluster is .74, within the cluster at the negative pole .66, while the average between-cluster correlation is −.65. These data provide additional verification of the bipolar nature of Factor *B*, and add emphasis to the statement just made regarding the general significance of the dimension defined by Factor *B*.

The bipolar designation "Idealism" vs. "Practicality" will serve to identify Factor *B*.

Second-Order Factor C: *Religiosity*

Variable 53.[11] Loading, .83. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's religious faith; rather than toward one's ideals of beauty, one's business organizations and associates, society as a whole.

Variable 40.[11] Loading, .81. If he lived in a small town and had more than enough income for his needs, he would prefer to use it to endow the church to which he belongs; rather than apply it productively to industrial development, give it to a university for the development of scientific research, or devote it to hospitals.

Variable 26.[11] Loading, .72. Taking the Bible as a whole, he would prefer to regard it as a spiritual revelation rather than from the point of view of its beautiful mythology and literary style.

Variable 9. Loading, .57. He should prefer to read the headlines "Protestant Leaders To Consult On Reconciliation" rather than "Great Improvements In Market Conditions."

Variable 13.[11] Loading, .55. He does *not* believe that the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first cause, cosmic purpose, or God behind it.

Variable 59.[11] Loading, .41. He would like to think of Leonardo da Vinci's "The Last Supper" as expressing the highest spiritual tendencies and emotions; rather than as one of the most priceless and irreplaceable pictures ever painted, in relation to Leonardo's versatility and its place in history or as a masterpiece of design.

---

[11] Both statement and sign of factor loading have been reversed.

Variable 55. Loading, .41. Florence Nightingale interests and attracts him more than Napoleon, Henry Ford, or Charles Darwin.

Variable 54.[11] Loading, .40. He believes that one should guide one's conduct according to, or develop one's chief loyalties toward, one's religious faith, one's business organizations and associates, or society as a whole; rather than toward one's ideal of beauty.

Variable 32. Loading, .39. He believes that good government should aim at more aid for the poor, sick, and old, the introduction of more ethical principles into its policies and diplomacy, the establishment of a position of prestige and respect among nations; rather than the development of manufacturing and trade.

Variable 42.[11] Loading, .39. When he goes to the theatre, he enjoys most, as a rule, plays that treat the lives of great men, plays with a theme of human suffering and love, or problem plays that argue consistently for some point of view; rather than ballet or similar imaginative performances.

Variable 24.[11] Loading, .39. He feels that the world would be a much better place if we took to heart the teaching, "Lay not up for yourselves treasures upon earth . . . but lay up for yourselves treasures in heaven, where neither moth nor rust doth corrupt, and where thieves do not break through nor steal."

Variable 16.[11] Loading, .38. At an exposition he prefers to see automobiles rather than scientific apparatus or chemical products.

Variable 21.[11] Loading, .36. He feels that the aim of the churches at the present time should be to convey spiritual worship and a sense of communion with the highest; rather than to bring out charitable tendencies and to urge people to think of the good of others.

Variable 49.[11] Loading, .36. During his next summer vacation he would prefer to stay in some secluded part of the country where he could appreciate fine scenery, go in for a local tennis or other athletic tournament, or get experience in some new line of business; rather than to write and publish an original biographical essay or article.

Variable 28.[11] Loading, .36. If he were to choose a topic on which to write an essay, he would choose the role of church-going in religion rather than the defects of our present educational systems.

Variable 18. Loading, .34. He would prefer to hear a series of lectures on the comparative development of the great religious faiths rather than on the comparative merits of the forms of government in Britain and in the United States.

Variable 11.[11] Loading, .34. When he visits a cathedral, he is more impressed by a prevading sense of reverence and worship than by the architectural features and stained glass.

Variable 10.[11] Loading, .31. He would prefer to read the story "Threat to Constitutional Government in Germany by a Dictator" rather than "Physicists Make Important Discoveries Concerning the Structure of the Atom."

From Table 6 it will be noted that Factor *C* has its heaviest saturation in Factors IV and VI, the "antireligious" and the "humanitarian" factors. Examination of the above listing of the particular items or variables which have the highest loading on Factor *C* indicates that with the exception of Variables 34 and 35, most of the variables having high loading on Factor *C* also have high loading on Factor IV and appear to be religious in nature. The relation of Variable 34 to religious evaluative tendency is fairly obvious. In the case of Variable 58, which has marginal loading, the alternative, "fundamentally spiritual in her attitude toward life," very probably accounts for its loading on Factor *C*.

While Factor *C* appears to be religious in nature, the exact nature of the religious evaluative component involved and the exact manner in which it differs from Factor IV is not directly evident.

It appeared plausible to the author that Factor *C* would disappear or reduce to a factor of rather narrow content if Factor II and other of the first-order factors were rotated in an appropriate manner. A technique for accomplishing these rotations without successive recomputation of and refactoring of the intercorrelations of the primaries is not, to the author's knowledge, available.

D. Discussion

In comparing the over-all results of

the factor analysis here reported with the purported content of the six scores for the Allport-Vernon test several points are evident. *Aesthetic, antireligious, theoretic,* and *social* factors resembling the original scores on the test were identified. Common factors with content having reasonably close resemblance to that of the *economic* and *political* scoring keys failed to appear. Items in these two evaluative scores are, however, prominent in the second-order general Factor *B*. Thus, it might be said that four of the evaluative scores give estimates of important factors resulting from the present analysis, and that the fifth and sixth may be combined to give an estimate of a single factor. *However, each of the groups of items in the four scoring keys which resemble factors resulting from this analysis was split into at least two factors.* For example, in addition to the general aesthetic factor, an "Interest in Fine Arts" primary trait was isolated; a factor labeled "Interest in Culture," involving a number of religious items, was isolated in addition to the antireligious factor; an "Interest in Physical Sciences" factor was obtained in addition to the theoretic factor. Thus, while the analysis resulted in factors having agreement with the scores of the test, it must also be stressed that the test scores are not factorially pure, but must be considered as an average of one factor accounting for a large proportion of its variance, and one or more additional but usually less important factors.[12]

In general, there was closer agreement with the scoring key of the test than with the results of Lurie's analysis. The identification of a general factor heavily involving both the political and economic evaluative tendencies was, however, a common finding of both analyses, which was inconsistent with the content of the scoring keys. The absence in Lurie's analysis of a general aesthetic factor was the most marked discrepancy between the results of the two analyses, although a reinterpretation of Lurie's results did suggest that one of the factors which Lurie isolated but did not identify showed appreciable resemblance to the general aesthetic factor of the present analysis. While there might seem to be considerable superficial resemblance between Lurie's social and theoretic factors and Factors VI and IX of the present analysis, close examination of the component variables loaded in the factors indicated few identical or high similar variables and little basis for assuming identity of these two pairs of factors. Of the several factors in the two analyses having approximately synonymous names, only the religious factors found in the two analyses may be regarded with assurance as closely resembling each other. In addition, there were a considerable number of primary evaluative tendencies isolated in the present analysis with some confidence, over and above those identified by Lurie.

The fact that Lurie constructed a test in which evaluative tendency was stated directly rather than in relation to other possible evaluative tendencies may account in part for these differences in results. In addition, although Lurie's test was intended to duplicate the content of the Allport-Vernon test, there was appreciable difference in content as well as in form. Lurie's practice of adding several items to obtain a single variable probably added to the factorial complexity of his variables, and may have been a still additional factor partly responsible for differences between the results of his analysis and those reported in the present paper.

The general nature of the primary traits as they were isolated and interpreted in this paper does not support Spranger's theory of evaluative types insofar, at least, as the test scores conformed to that theory. An aspect of the interpretation of a number of the factors isolated is pertinent to this point and will bear emphasis in its own right. In interpreting several of the factors resulting from the present analysis, it appeared that the beliefs and evaluations stated in variables with high loadings were most readily interpretable as characterizing a particular social group or institution. It seemed quite clear, for example, that a high or low score on Factor IV (antireligious) depends upon disagreement or agreement with the conventional religious beliefs of the Christian church. Factor VII seemed readily interpretable as an "Interest in Science" factor. If we regard science as an institution, differences between individuals scoring high and low on Factor VII can be ascribed to differences in their interest in or evaluation of science as an institution. Again, Factor X ("Rugged Individualism") appeared due to individual differences in acceptance or rejection of the evaluations inherent in the American tradition of rugged individualism.

---

[12] This comparison of the factor content with the content of the scoring keys is subject to some qualification, since two alternatives of each of fifteen four-choice items in Part II of the Allport-Vernon test were not included in the analysis.

Factor *B*, the most prominent in terms of the proportion of common variance accounted for, was interpreted as representing the divergent viewpoints of the idealist and the practical man. It was postulated further that the apparent bipolar nature of the factor is related to formation of opposing "we groups" and resulting conflict between them. While in this instance the social structures postulated are not formal or institutionalized, it is believed that this interpretation of Factor *B*, if correct, supports the generalization stated above regarding the importance of social groups in understanding and interpreting the obtained factor structure.

VI. SUMMARY

The Allport-Vernon test, "A Study of Values," was administered to 200 college students. Tetrachoric intercorrelations of 60 of the items were computed, and eleven centroid factors were extracted and rotated to an oblique simple structure. Three second-order general factors were determined from the intercorrelations of the oblique factors, and the loadings of the tests on the second-order factors computed.

The first-order factors described below were identified with varying degree of confidence.

Factor I. *General Aesthetic Interest*

Factor I is a general aesthetic factor which corresponds rather closely to the aesthetic measure obtained through the scoring key of the Test of Values.

Factor II. *Interest in the Fine Arts*

Factor II is less well established and less general than Factor I. It appears to involve group-held evaluations of the artistically inclined. The content seems restricted to the fine arts.

Factor III. *Belief in "Culture"*

Persons scoring high on Factor III can be characterized by a desire to appear "cultured." Their evaluations appear to have been determined by processes that might be described as intuitive or mystical. The content of items having positive correlation with Factor III tends not to involve apparent economic, biological, or social needs.

Factor IV. *Antireligious Evaluative Tendency*

Factor IV appeared to be definitely identifiable as antireligious in nature. Differences in reaction toward the conventional creed or indoctrination of the Christian church as an institution were postulated as the basis for the variance accounted for by Factor IV.

Factor V. *Antiaggression*

Factor V apparently has to do with individual differences in interests or evaluations relating to conflict, possibly to armed conflict in particular.

Factor VI. Humanitarian Tendency

The items of the Test of Values scored as social were well represented among those having high loading on Factor VI. Since the items highly loaded on Factor VI were largely concerned with social issues, the interpretation of Factor VI simply as "social" was not adequate. The personal or face-to-face social interests which appear to have been postulated in many of the items scored as social did not appear to be prominent in the content with high loading on Factor VI. A sympathetic or humanitarian rather than a rational basis for such interests in social issues seemed implied by the variables with high loading on Factor VI.

Factor VII. *Interest in Science*

Factor VII seemed definitely related to high evaluation of, or interest in, the physical sciences.

Factor VIII. *Tendency toward Liberalism*

Factor VIII resembled Factor VI in many respects. Persons high on Factor VIII, however, seemed characterized by a liberal viewpoint on social issues which seemed to be intellectual rather than emotional in origin.

Factor IX. *Theoretic Interest*

Factor IX seemed to correspond closely to the conception of theoretic interest indicated by Spranger's definition as evidenced in the scoring key for the Test of Values. It should be noted, though, that items dealing with social issues and with interest in science were not highly loaded on Factor IX.

Factor X. "Rugged Individualism"

Tentatively, individual differences in acceptance or rejection of the American tradition of "rugged individualism" were postulated as the source of the variance accounted for by Factor X.

The meaningful interpretation of Factor XI was attempted. It was assumed to be residual in nature.

Of the three second-order factors isolated, Factor *A* did not account for a large proportion of the variance of the original matrix and was not interpreted, while Factor *C* seemed too similar to first-order Factor IV to add to the interpretation of the primary evaluative tendencies as described by the first-order factors. Factor *B*, on the other hand, appeared to be the most important source of variance in evaluative tendencies revealed by the analysis. Factor *B* was interpreted as a basic source of individual differences in evaluative tendencies which differentiated between *idealism* on the positive pole and *practicality* on the negative pole. The term "idealism" was used to describe an interest in, and high evaluation of, concepts and beliefs concerned with improvement of the social structure and the general manner of living of human beings. Practicality implied acceptance of "things as they are" with high evaluation of activities, concepts, and attitudes having to do with effective action within this framework.

REFERENCES

1. Allport, G. W., & Vernon, P. E. *A study of values.* Boston: Houghton Mifflin, 1931.
2. Duffy, Elizabeth, & Crissy, W. J. E. Evaluative attitudes as related to vocational interests and academic achievement. *J. abnorm. soc. Psychol.,* 1940, 35, 226-245.
3. Guilford, J. P. Personality factors S, E, and M and their measurement. *J. Psychol.,* 1936, 2, 109-127.
4. Mosier, C. I. A factor analysis of certain neurotic tendencies. *Psychometrika,* 1937, 2, 263-287.
5. Lurie, W. A. A study of Spranger's value types by the method of factor analysis. *J. soc. Psychol.,* 1937, 8, 17-37.
6. Strong, E. K. *Manual for Vocational Interest Blank for Men.* Stanford Univer.: Stanford Univer. Press, 1945.
7. Vernon, P. E., & Allport, G. W. A test for personal values. *J. abnorm. soc. Psychol.,* 1931, 26, 231-248.

Accepted for publication, March 13, 1952.

Vol. 66, No. 17 Whole No. 349, 1952

Psychological Monographs: General and Applied

Selected Personality Variables and the Learning Process

Eugene L. Gaier[1]
University of Illinois

I. INTRODUCTION

Changes in overt behavior or verbal performance following practice with some verbal or motor task have generally been considered as evidence of learning. These changes of behavior have been primarily investigated from two approaches: *(a)* studying the resultant changed behavior; and *(b)* studying the processes leading to the observed changes in behavior. While both of these approaches are essential, they must be related to a third variable: the nature of the individual involved in the task. It is in the light of the needed research in the relationship between the learning situation and the personality of the learner that the present study has been initiated.

We are here making the initial assumption that certain relatively stable personality characteristics exist which are reflected in the individual's handling of the learning situation. Freud asserted that the thinking processes were conveyors of underlying feelings, and that there seemed to be a definite connection between the thought processes and the other processes occurring simultaneously within the individual. This position has been taken for the present investigation, which aims to demonstrate the relationship between the individual's kinds of thinking and selected personality characteristics.

In an attempt to get at the manner in which the individual's personality characteristics are manifested in the learning situation, the present study is designed to investigate the relationships between the recalled verbalizations of associations experienced in a series of learning situations over a period of four months by a group of college students in the University of Chicago, and personality characteristics as inferred from test protocols of the Rorschach ink blot test. The specific questions we were here interested in investigating were:

1. What is the relationship between selected personality characteristics (viz., anxiety, rigidity, negativism) and the kinds of thoughts a group of college students have in a series of learning situations?

2. What is the relationship between selected personality characteristics and products of learning, where such products can be inferred from test performance?

Many investigations have attempted to

---

[1] This study is a revision of a dissertation submitted in 1951 to the faculty of the Department of Psychology in the Division of Social Sciences of the University of Chicago in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

To Dr. Benjamin S. Bloom, under whose direction this study evolved, the writer is indebted for helpful suggestions, critical judgments, and encouragement. The writer also wishes to express appreciation to Dr. Ernest A. Haggard and Dr. Donald W. Fiske for their stimulating interest and counsel.

formulate concepts linking levels of individual functioning conceived as parts of one process. Research in social psychology and perception has indicated clear relationships between perception and personality. The present study is a departure from these lines insofar as mental processes going beyond immediate perceptions are studied objectively in an effort to understand how they relate to some general and stable personality characteristics. Here we are attempting to investigate the relationship between the more highly cognitive ability of problem solving and personality characteristics.

II. THE METHOD OF STIMULATED RECALL

A. Background

The class situation is a process consisting of overt, observable behaviors or activities and nonovert behaviors or private experiences of both the teacher and the students which defy observation. Until the present, the major source of evidence for studies of classroom learning and procedures has been the investigation of the overt behaviors. Where mental processes of students have been considered, they have been assumed or inferred from the overt behaviors. Those who propose theories about learning acknowledge that some kinds of mental activities, as well as physical activities, are involved in the learning process. The physical activities are presumably accompanied by certain conscious and unconscious activities. The techniques for obtaining evidence of learning, however, are limited by current methods to data of physical activities, i.e., overt, observable behaviors. The conscious and unconscious mental activities regarded important for learning are less readily observable.

The extent to which mental processes can be inferred from overt behavior is limited since specific behaviors are not necessarily accompanied by the same mental processes in different individuals. Any attempt to infer the mental processes from observations of overt behavior is subject to much error, since a variety of mental processes may lead to the same end product. And unless additional evidence is gained about the *processes* involved, the investigator cannot be certain of his inferences.

In the attempt to obtain evidence of the thoughts occurring in a learning situation, subjects have been asked to observe their thinking while participating in an activity and to report this upon the completion of the task. Titchener (19) had his subjects think aloud while engaged in problem solving. Often, the process of observing one's own thinking affected the very behaviors the experimenter was interested in investigating. The attempts to obtain evidence of the conscious processes by asking participants to report and record their thoughts during a given situation have not been very successful. Bloom (1) noted that it was not feasible to have psychologically naive subjects both participate in an activity and simultaneously observe their participation. On the other hand, if psychologically trained subjects are to give introspective accounts as they participate in an activity, their training and psychological sophistication would render them poor subjects for investigating many complex social situations such as learning a subject at the college freshman level.

B. The Method of Stimulated Recall

The method of Stimulated Recall, introduced by Bloom (1), is a promising technique for attempting to get at the conscious thoughts of the individual. Basically, this is a technique which attempts to get at the conscious thoughts, ideas, and experiences an individual has had in a learning situation. This method is based on the assumption that given enough of the original situational material (i.e., cues from the class session), the individual will be stimulated to recall not only the original overt activity (i.e., the verbal and physical behaviors of the class situation) but his own private experiences as well. Sound recordings made of the class discussion are played back to the participant in an attempt to stimulate recall of his thoughts. It is these recalled and reported experiences that become the data obtained with this technique. No special training is required for the subjects to recall and report their thinking. This participant-observer method, unlike introspection, permits full, unhampered, and free participation of the subjects during the original situation. Thus, the individual becomes both the participant and the observer.

The method of stimulated recall lends itself not only to providing reports of the mental activity occurring concurrently with certain overt activities, but also in making comparisons of all members of any situation at a defined period. This method is applicable to group situations in which there are many participants and where considerable overt verbal activity occurs. It is assumed that the sound record will provide sufficient cues of the original situation for the participant to recall his experiences. Within a 48-hour period following the original situation, each participant of the group is individually interviewed and the sound record is played back to him. At specific, preselected points in the discussion, called *Critical Points,* the recording is stopped and the subject is asked to recall and report the thoughts he experienced in the situation at this particular portion. The reports given at the stops of the recording provide evidence of the conscious processes.

III. THE COLLECTION OF THE DATA

A. THE COLLEGE AS A LEARNING CONTEXT

Education in the College of the University of Chicago is based on the idea that the students take general courses in several major areas, with work presented principally through small discussion classes. Supervision of the student's work is considerably less than in high school. A student's standing is measured solely by performance on comprehensive examinations given at the end of the school year, rather than by such indices as class attendance, teacher grades, or teacher-constructed tests given during the year.

At entrance, records of previous academic work are not considered in determining the student's course placement. The need for standard measures of competence, in view of the variety of backgrounds of the students, has led to the placement-examination program. While school records provide some indication of previous training, a list of credits is often an inadequate basis for determining the point at which the student is prepared to begin his studies in the College. Some entering students have completed but ten grades of schooling, while others have had several semesters of college work. The development of these placement examinations has served to indicate the level of courses and the number of courses to be taken by the applicant, in terms of his particular academic needs.

At the outset, it must be pointed out that the subjects used in this investigation, if representative of the University of Chicago College entrants, are not representative of the *usual college fresh-*

man. As noted by Cronbach (4), they are intellectually extremely superior and have special motivations which cause them to seek out the Chicago program. The context in which the entering College student finds himself is one of intellectually superior and aggressive students with high motivation and a keen sense of competition. Every attempt is made to meet the individual's intellectual needs and abilities, both in the initial screening of applicants and in the placement of the admitted students in the course. It is within this type of context that the present subjects were drawn and the study was conducted.

B. The Subjects

The subjects for this investigation consisted of members of one intact Social Science I class of 11 students, registered for the first time in the College of the University of Chicago. These 3 female and 8 male students ranged in age from 15 to 22 years. These subjects were simultaneously registered in either Mathematics I or Humanities I, in addition to the Social Science I course. This class in Social Science, selected as the parent course, was chosen because of the degree of involvement afforded the student by the subject matter presented therein. The range of materials covered contains strong implications for everyday life beyond the college situation, serving to arouse a greater sphere of thinking related to personal values and attitudes. The topics covered succeed in involving the students in a manner not usually possible with subject matter such as mathematics, chemistry, or a foreign language.

C. Procedures in Data Collection

We were interested in determining the manner in which selected personality characteristics are manifested and employed in varying types of learning situations. The present investigation is based on three types of data: *(a)* the conscious thoughts or ideas experienced in the classroom and reported in individual interviews over a period of four months; *(b)* ratings of Rorschach test protocols; and *(c)* performance results on achievement and aptitude tests given college students.

1. *The Conscious Ideas*

In order to familiarize each class with the presence of the recording equipment and the investigator, preliminary class recordings were made. Prior to the initial recording, each of the 11 members from the class selected for study was individually interviewed in an attempt to secure his cooperation for this study. Each student was informed that a series of individual interviews would be a necessary part of the investigation, and would extend over a period of 4 months. At the time of the initial contact interview, the students were not informed as to the nature of the interviews that would be conducted. Each person interviewed consented to participate in the investigation.

Each class studied was recorded for the full 50-minute session. From the class discussion, statements made by students or read from the text which appeared to have the most group involvement were selected to serve as critical points during the interview sessions to stimulate the recall of the original thoughts. An attempt was made to select those class materials that would lend themselves to maximum personal meaning and involvement on the part of the students, as well as to include many types of classroom verbalizations in the 10 to 15 statements or phrases selected from each class. No method, however, was devised for a random sampling of these statements.[2]

Within a 48-hour period of the class recording, each subject was individually interviewed. At the beginning of the first interview, the following directions were given each subject:

> The College has been interested in studying the effects of teaching and class procedures used here. In the past, recordings of the class sessions have been made and studied in great detail to enable the instructors to become more sensitive to their own teaching methods. Now, in addition, we are interested in investigating the kinds of things that students think about during the class discussion.
>
> The method that we are making use of here is to play back the recorded session of the class in which you participated this morning [yesterday or this afternoon]. The record which I have here will be played back to you in order to help you recall the thoughts and ideas you had while the class was going on. This is similar to asking you to recall your tenth or twelfth birthday if I should tell you about the gifts you received or the songs that the guests sang to you.
>
> It is unlikely that you will be able to recall very much at first, but as the recording of this class is played back to you, and as you get more of the feeling for the class session, you should be able to recall more and more of your classroom thoughts.
>
> As the recording is played back to you, I shall stop it from time to time at certain points, and I shall ask you to recall what you were thinking about during the original situation. Feel free to recall and relate any experiences or ideas you had in the class situation, no matter how trivial or irrelevant you may feel them to be.
>
> If you recall any ideas or experiences you had during selections of the recording at which I do not stop, tell me, and I shall stop the recorder. Anything that you tell me will be recorded verbatim. This information will be held in confidence, and will not be shown to any of your instructors or used in any way to evaluate your work here in the College. Have you any questions before we begin?

The verbatim reports of the experiences and ideas recalled at the given critical points of the subjects provided the major portions of the data for the present investigation. Henceforth, these recorded reports will be referred to as the *conscious experiences or conscious ideas.* A sample of 107 overt critical points was used at which the subjects were questioned for their recall of class experiences. This total included 77 preselected points, and 30 points at which the subjects volunteered their experiences. A total of 2166 conscious idea units was reported. Of this amount, 1937 units were actually experienced in the class situation, and 229 units were volunteered as a function of the interview situation.

2. *The Projective Test*

Each of the 11 subjects was individually administered the Rorschach ink blot test during the last month of the investigation period. Test administration was delayed until all of the interviews had been secured, since it was felt that the unstructured nature of the Rorschach, plus the common misunderstandings of the nature of the ink blot series, might arouse resistance and possibly serve to terminate the interviews, vitiating our original design. The ratings of various dimensions of personality inferred from the Rorschach constitute our second source of data.

3. *Test Situational Data*

Entrance Tests. Each College applicant is given a series of entrance tests consisting of the *American Council on Education Psychological Test* (ACE), a *Reading*, and a *Writing Achievement* test.

---

[2] The proportions of the various overt verbal behaviors selected for presentation to the subjects, however, were approximately the same as the random sample selected from the data for the entire study, determined by selecting every fourth statement throughout the entire typescripts of the recordings.

Evidence gathered over a period of years has indicated the value of these tests for predicting success in the College; and the test scores constitute part of the information used in admitting applicants.

Placement Examinations. Upon being admitted to the College, each student is given a series of placement examinations. These tests are planned to measure skills in reading, interpretation, and the solving of problems, as well as general subject-matter knowledge. The results of this series of tests, covering all areas of college work, are used as an index for placing students in the courses geared to individual needs and abilities. On the basis of placement-test achievement, students may be classified in four groups in each subject field:

a. Those who have competence equivalent to that called for by a particular comprehensive examination and are excused, therefore, from the examination.

b. Those who have mastered a certain portion of the knowledge and skills covered in a comprehensive examination and are advised, therefore, to take only a part of the course offered in preparation of the examination.

c. Those who have not acquired sufficient understanding and skills to pass a particular comprehensive examination without first taking the entire course in preparation for the examination.

d. Those who are likely to fail in their examination unless they receive special training in addition to the preparation regularly offered in the course.

The performance of all of the 11 subjects on the examination fell into Category *c.* Because they had not acquired sufficient understanding and skills to pass the placement examination (i.e., and go on to the next course in the Social Science sequence), this is assumed relatively constant for the group.

College Achievement. The use of the ACE provides some index of potential ability the student brings with him upon entrance, as well as some basis for the prediction of his subsequent school success. The placement examinations provide a measure of achievement *at the time of entrance.* The third type of data obtained was individual performance on the Social Science I Comprehensive Examination. This comprehensive examination, constructed by a central committee, is given at the end of the year. The examiners work closely with the instructors in the preparation of the test, but the final form of the examination is prepared by the central committee of examiners. Final grading is made without knowledge of the identity of the writer of the paper, to minimize bias.

The comprehensive examinations are composed of a series of parts designed to tap a variety of behaviors and skills that meet the objectives of the College. It is recognized that the single grade given the student for achievement on the examination does not adequately describe his performance, since it is unlikely that performance is consistently at the same level throughout the various processes tested. Some parts of the examination represent subject-matter divisions and facts, while others call for the ability to analyze and organize new or familiar subject matter. The areas of the Social Science I examination are fully presented in Part IV.

In summary, three kinds of test situational data were collected: one of aptitude (ACE) and two of achievement. One form of achievement is based on work previous to entering the College, while the comprehensive examination is a measure used after subjecting the student to a series of specific learning materials in a systematic fashion.

4. *The Rorschach Results*

In our selection of the kinds of relatively stable personality characteristics that would become manifest in the learning situations, two criteria were employed:

a. Can the personality characteristics, assumed to be stable, be inferred from the Rorschach records with sufficient reliability (at the 5 per cent level) by two judges?

b. Can the kinds of personality characteristics previously studied and reported in the literature become manifested in the kinds of data obtained here?

Among the personality characteristics studied in the literature have been (*a*) free floating anxiety; (*b*) negativism; (*c*) rigidity; and (*d*) aggression. It seemed of interest to attempt to relate these characteristics as evidenced on a projective test, with the indices of intelligence and accomplishment used here (ACE, achievement tests, and grades).

Rank-order correlations were computed between ratings made independently by two Rorschach experts. Use was made only of those ratings whose reliability coefficient[3] was .601 or over (significant at the 5 per cent level). The correlation between negativism and aggressiveness (.67) was high enough to lead us to believe that one may be but a facet of the other, and separate treatment would not add to the interpretation of the data. The ratings made by the two experts were pooled, and were then assigned each subject.

5. *The Conscious Behaviors*

A system of classification of the recalled conscious experiences was developed for the present study. Each separate idea, in many instances a clause or phrase, served as the unit for classification. This unit of division, here called the *idea unit,* was selected, since it appeared that a classification division based on the number of words spoken, or on all materials expressed at one critical point, would serve to obscure the richness of the obtained data. The need for this particular system[4] of classification becomes clear when material such as the following is expressed at one critical point:[5]

> Donna:/ When I first heard of the survival of the fittest in biology class in high school and the Darwinian theory, well, I was thinking of this./ Thought I would like to learn more about all of this./ This expression surprised me, for I never heard about the Judeo-Christian tradition. People usually just say "Christian Tradition."/ I was laughing at Jack's expression, "Someone tried to hit me over the head." I wasn't thinking about it. It just amused me./ I agreed with Mr. X— (the instructor)./ I wondered what the word "progeny" meant, and where he got it./

Where more than one discrete thought was reported at a particular point in the original situation, each idea was classified as a separate unit according to the method of division adopted here. It was the rare critical point that stimulated the recall of only one idea unit.

In addition to the ideas actually experienced in the actual class situation, the following kinds of thoughts and information were reported:[6]

1. Ideas experienced immediately preceding the critical-point stop. This was usually presented by the subject to give continuity to the ideas expressed during the interview.

---

[3] Spearman rank-order method.

[4] At best, the classification of any qualitative material is subject to some distortion once it has been removed from its original context. The classification title is but a convenient symbolic shorthand that lends itself to a more facile treatment of the data than would be the case if it were left in its original context.

[5] The unit division is indicated by a diagonal bar.

[6] These were not used in the analysis of the data reported here, however.

2. Explanatory remarks made in reporting the actually experienced ideas. These were usually given for the benefit of the interviewer. These ideas were recorded in a special way to take care of their source of stimulation as attributable to the interview situation itself.

3. New ideas and associations which the individual experienced during the interview situation. These comments usually arose as explanations, attitudes, or individual perceptions of classroom situations, the instructor or classmates' behavior. Most common, however, were affective feelings of the self, and progress the individual felt that he had been making.

4. Reports of the overt behavior or conscious experiences following those at the critical point.

Source of Thought. Each idea unit was classified according to the stimulus or *objective source* of the idea, as inferred by the judges from the subject's report. This classification dealing with the source of the idea units will, henceforth, be referred to as the *Source of Thought.*

A tenfold classification was developed in an attempt to describe the sources of ideas during the actual classroom process. This classification may be broken down into the degree of relevancy and desirability for classroom thinking:

1. The highest degree of concentration in the classroom would be problem-oriented thinking concerned with the topics of the educational experience. Here, we would expect the source of thought to be the ideas directly under consideration. Included here were the categories: (*a*) *ideas under consideration;* and (*b*) *a related idea.*

2. Just beyond the immediate subject matter may be found kinds of thinking not directly stimulated by the problems discussed. Here the ideas stem from the course materials in the form of reading in the text, looking through notes, or working problems. Included here was: (*c*) *activity related to the class or subject independent of the ideas directly under consideration.*

3. Not related to the class material, but still *confined to the class situation,* is the student's thinking about people in the classroom. The instructor, classmates, and the self are all included here in the categories labeled: (*d*) *self;* (*e*) *teacher;* and (*f*) *another student in the class.*

4. Thoughts stimulated by stimuli within the class situation may include: (*g*) *objects within sight or hearing;* and (*h*) *classroom situation and/or process.*

5. The student may introduce stimuli external to the class process and surroundings. These are signs of distraction and diversion in the class experience and include: (*i*) *someone not in the classroom;* and (*j*) *some concrete object or situation not in the immediate environment.*

6. For units not appropriate to the above ten areas, a special category entitled *unclassified* was employed.

Discomfort-Relief Quotient. A modification of the Discomfort-Relief Quotient (DRQ) developed by Dollard and Mowrer (5) was employed for classifying this dimension of the experienced ideas. As a measure of progress in case work, these authors obtained the DRQ by having judges divide the transcripts of the protocols into clauses and thought units. Each unit was classified as an expression of discomfort, relief, or neutral affective tone, and the tension index was in the form of:

$$\text{DRQ} = \frac{\text{Number of Units of Tension}}{\text{No. of Tension Units} + \text{No. of Relief Units}}$$

Methods of classification of units set up by sentences, paragraphs, or pages were reported by these authors. But the idea unit, as adopted for the present study, was reported as being most reliable.

On this basis, the following classification was employed here:

1. Direct expression of relief, lowered tension, or satisfaction.

2. Direct expression of discomfort, tension, or annoyance.

Clearly, the use of only statements of direct expressions of satisfaction or dissatisfaction would result in discarding much of the data. Consequently, two other categories were developed to include units where no explicit expression of affect existed:

3. Inferred relief, lowered tension, or satisfaction.

4. Inferred discomfort, tension, or annoyance.

To account for units without affect or positive or negative tone, two additional categories were set up:

5. Neutral
6. Ambiguous and/or ambivalent.

Orientation of Thought. While many individuals may be aroused by the same stimulus (our *Source of Thought* classification), each individual may handle and treat these stimuli in differing ways. Some classification seemed desirable here to take into account the differing characteristics and treatments and orientations of the idea units reported. In this classification, henceforth to be referred to as the *Orientation of Thought,* no attempt was made to judge the quality or the appropriateness of the reported experiences. The basis for describing any content was to be found in the idea unit.

In brief, a classification believed to lend itself to describing the orientation of class experiences was developed. Four kinds of general content trends were evidenced in the obtained data:

1. Active thinking or problem solving, indicating an actual awareness and comprehension or evaluation of the topics discussed. Included here were the categories *(a) active participation; (b) evaluation;* and *(c) comprehension.*
2. Passive thinking about class topics, indicating an awareness of the problem but not an active consideration. This category was labeled *(d) passive participation.*
3. Under the broad heading of *Diversions* were included the reports of emotional reactions and attitudes about the self, other people, and class materials. Categories contained here were: *(a) individuals and their characteristics; (b) ideas and words;* and *(c) concern about the self.*
4. Reports of class activities and conditions. This category was labeled *(d) other activities.*

In addition, the classification of III. *No recall* was included to account for the failure by the subject to remember the ideas experienced in the class at the preselected critical point.

Each of the first eight general categories of the above ninefold classification was subdivided. (See Appendix for a complete description.) Each idea unit, taken from the written typescript of the reported experiences, was classified as to *one* kind of orientation of thought.

In the ordering of qualitative data, the danger of the classifier's own bias exists; and lack of awareness of this bias on the part of the investigator may serve to distort the results obtained. At the outset of this investigation, in order to minimize classifier distortion, it seemed advisable to adopt a systematic bias in classifying the idea units both for *Source of Thought* and *Orientation of Thought.* The bias adopted was in the direction of the desirable learning in the college situation. In the ordering of the Units, a hierarchy extending from class-oriented thinking to individual-centered experiences was set up. Wherever a choice was possible between problem-oriented and self-oriented thinking, preference was in the direction of the former.

6. *Objectivity of Classification*

The ability of other judges, in addition to the investigator, to learn the present classification systems and to apply them in the same fashion was the basis for the criterion of objectivity used in evaluating the categories here developed. All interview protocols were divided into idea units independently by the investigator and one judge to obtain an index of objectivity of idea-unit division. In addition, all of the protocols were classified by two judges plus the investigator for the *Source of Thought, DRQ,* and *Orientation of Thought.*

A practice session was held with each judge during which a random sample of idea units was coded. Differences in interpretation were then discussed in order to establish a common frame of reference among the judges and the investigator. Pearson product-moment correlations[7] were computed between the classifications of the two judges.

Agreement of Content Unit Division. All interview protocols were divided on the basis of the idea unit. Judgments were to be based solely on the completeness of an idea per se. *The agreement on unit division obtained between the judge and the investigator was 98 per cent.* For the remaining 2 per cent of units in which there was disagreement, the judges met to reconcile the differences. The high percentage of agreement obtained indicated sufficient objectivity to allow us to proceed to the other classifications.

Objectivity of the DRQ Category. Eleven interview protocols from one class were given to three judges with comparable psychological training, to make judgments of the modified *DRQ*. Two of the judges were instructed to make their judgments on the basis of deliberate exclusion of the possible psychological significance or implications of the responses. Judgments were to be made on whether the individual actually said or implied satisfaction, tension, etc. The correlations between judges and the investigator, based on 277 idea units, were as follows:

$$^{r}E \text{ \& Judge } A = .90$$
$$^{r}E \text{ \& Judge } B = .89$$

Judge *C* was given equal familiarity with the classification scheme but with one further instruction: to make inferences, with attempts to get at the psychological significance of the expressed and inferred hedonic tone. Correlated with the classification made by the investigator on the same basis, the obtained value of *r* was .69. While lower in magnitude than the coefficients obtained when the idea units were classified on actual report, as adopted here, this value does indicate that the objectivity of both types of classification is high.

Objectivity of the Orientation-of-Thought Categories. The same 11 interview protocols were classified by one judge and the investigator, and objectivity of classification in broad categories was measured by the Pearson *r*. A value of .94 was obtained. A reliability coefficient of .64 was obtained for classification of the subcategories of the nine major divisions. These values indicate that both classification in broad and subcategories of the orientation of thought was sufficiently objective for use here.

Objectivity of the Source-of-Thought Category. A sample of 277 idea units was classified by one judge and the investigator for source of thought. The obtained reliability coefficient was .95. All of the above correlation values are significant, and we may conclude that the classification system developed and used here is of sufficient reliability and objectivity to serve as a basis for testing our hypotheses.

---

[7] It is recognized that the assumptions (e.g., continuous variable, homogeneity, homoscedasticity, numerous arrays) for the computation of the Pearson *r* were not met here. The most appropriate measure here would have been use of the contingency coefficient. Since many of the *theoretical* cell frequencies were less than the value of five, the value assumed necessary by Edwards (6) for legitimate use of this method, contingency coefficients were not used. The obtained values of the Pearson *r* do indicate, however, a high level of agreement between the judges as to the objectivity of the classification categories developed.

IV. ANXIETY IN THE LEARNING SITUATION

A. Anxiety as Measured by the Rorschach

The Rorschach record is capable of providing two types of indices of anxiety. With a well-systematized type of anxiety, i.e., what May (13) has termed *neurotic anxiety,* the individual has already produced a neurotic defense system, evidenced by marked changes of personality structure, resulting in both compulsive and hysterical personality characteristics. Klopfer and Kelley (11), however, suggest that this type of anxiety is less likely to become manifest on the Rorschach protocols. They write that it is the reactive depression or anxiety state, referred to as an unsystematized anxiety, that affects the record more easily. Carried one step further, it is an inner tension of a generalized "free-floating" nature, and represents an ineffective effort (as inferred from the record itself) to intellectualize the anxiety. Neurotic anxiety, moreover, involves the mechanism of repression, and differs from this free-floating or normal anxiety by its symptomatic nature. In May's words:

> Normal anxiety [is] that which is used for a constructive solution to the problem which causes the anxiety, and neurotic anxiety [is] that which results in defense from avoidance of that problem (13, p. 199).

Following May, anxiety is here conceived as a pervasive inner tension motivating the organism to protect itself against anticipated psychic pain. Since individuals differ with respect to ego strength or inner psychological security, it is reasonable to suppose that they also differ with respect to anxiety level or readiness. Correlatively, there may exist individual differences in the amounts of anxiety which can be handled without resorting to maladaptive responses. If anxiety exceeds the adaptive threshold, then rigid, *emergency* behavior may result. The individual's response latency, and thus his ability to perform noegenetically, is reduced, with the result that he is forced back upon familiar, habitual responses inappropriate in the given situation.

From these considerations, we may derive our two hypotheses: Hypothesis I: *If the anxiety is above individual threshold, it leads to an impairment in the ability to improvise in an unstructured and/or new situation.* This results in stereotyped, habitual, and familiar approaches that may be maladaptive in the situation. Secondly, excessive tension may so interfere with the perceptual processes as to considerably reduce problem-solving efficiency. The necessity for responding immediately may affect both the perception of the stimulus situation and the adequacy of the responses. This inability to be problem oriented may thus arouse concern for one's own psychological well-being, with increasing concern about self-adequacy for meeting the situation. From this we come to Hypothesis II: *If anxiety is above individual threshold, the individual becomes self-engrossed and concerned with his personal adequacy.*

To obtain an index of individual anxiety-readiness level, two Rorschach experts were asked to rank the 11 protocols in terms of this threshold. Individuals with a high readiness level, it was postulated, would be more sensitized to situational factors, and would be thrown into a disturbed state on less provocation than individuals with low anxiety readiness. The rank-order reliability of the two ratings (.80) was significant at beyond the 1 per cent level with 9 df. After the 11 subjects had been assigned pooled ratings on the basis of these judgments, they were divided into three groups taken as having low, average, and high anxiety readiness.

B. The Class Situation

Anxiety as a constructive agent accompanies adaptive striving by the total personality. It is conceivable that variation in classroom methods of treating the course materials may tap different levels of individual anxiety reactions. The lecture class, for example, represents a clearly defined stimulus situation in which the student need only sit and listen; he is not required to submit verbal evidence of comprehension.

The discussion classes in the College, on the other hand, present the student with a very different situation. Here the student is expected to prepare in advance, and to participate in discussing and solving the problems raised. There is the constant likelihood of being called upon to offer interpretations of ideas and concepts. It is important for the student to attack the problems correctly, since his treatment of any question is available for all class members to judge and evaluate. Motivation is ordinarily supplied by some form of tension or mild anxiety, and it may be assumed that the anxiety-relief attendant upon correct responses will constitute a positive reinforcement for the individual. This relief, however, is usually only temporary, for the anxiety may be rearoused by the anticipation of being called again where he might fare badly, and accordingly suffer deflation of self-esteem in the eyes of his classmates.

C. Evidences of Anxiety in the Test Situation

1. *Aptitude Tests.* The ACE provides two scores for each student: Linguistic (L) and Quantitative (Q). The L score represents knowledge of vocabulary and the ability to reason with words; the Q score measures skill and speed in arithmetical computation, numerical reasoning, and rather high-level nonnumerical reasoning. The Reading test given each College applicant calls for the use of verbal skills and inductive reasoning. The Writing test, more structured than the Reading, calls for the correction of poor grammar or errors in other aspects of the mechanics of writing. In the Reading test particularly, the student is faced with a situation calling for improvisation under the pressures of situational anxiety. The obtained correlations are presented in Table 1.

TABLE 1

RANK-ORDER CORRELATIONS BETWEEN ENTRANCE TEST PERFORMANCE AND ANXIETY READINESS, RIGIDITY, AND NEGATIVISM

Variable	Anxiety	Rigidity	Negativism
ACE: L Score..	−.37	.23	.20
ACE: Q Score..	−.39	.31	.16
Writing Test ...	−.40	.11	.19
Reading Test...	−.66*	−.16	−.04

* Significant at the 5 per cent level of confidence.

The values in Table 1 tend to support our first hypothesis: namely, that the ability to improvise with new materials decreases as the individual's anxiety-readiness level increases. The causation is probably circular, for poor ability may lead to anxiety as well as vice versa.

2. *The Comprehensive Examination.* Some parts of the Social Science examination represent subject-matter divisions while other parts are more specifically functional, in the sense that they tap relatively specific abilities and skills developed in the course. The parts range from the recognition-recall of factual material to the ability to apply methods

developed in the course to the handling of new materials. It should be noted that the student's course grade is determined solely by his performance on this comprehensive examination. Thus, his future in the College, as well as chances for scholarships and other awards, are all dependent on the results of these tests.

The Social Science examination in 1950 consisted of eight parts in dealing with the topics covered. These have been arranged here in four areas, each area designed to test the different types of required cognitive processes fostered by the College program. These include:

Part I. Recall through recognition of either (*a*) facts and generalizations about American history, or (*b*) the general nature of an author's position.

Part II. Analyze and interpret, either the papers read in the course or similar papers. This includes the ability to (*a*) identify the main argument of the presentation or discussion, or (*b*) recognize the kind of argument for policy of general views expressed by the author.

Part III. Compare authors, i.e., identify similarities and differences (*a*) as to how issues are defined, or (*b*) as to arguments and conclusions.

Part IV. Express in writing (*a*) simple analysis of any historical period studied; (*b*) the student's critical judgment on a problem relative to those dealt with in the course.

Considering the stressfulness of this test situation, we would predict from our first hypothesis that problems containing unfamiliar subject material would be less adequately handled than structured and familiar problems. The correlations between anxiety-readiness level and performance on the four areas of the Comprehensive are shown in Table 2.

TABLE 2

RANK-ORDER CORRELATIONS BETWEEN ANXIETY READINESS, RIGIDITY, AND NEGATIVISM AND THE FOUR PROCESSES TESTED ON THE SOCIAL SCIENCE I COMPREHENSIVE

Part	Process	Anxiety	Rigidity	Negativism
I	Recall through identification	.34	.73*	.03
II	Analyze and interpret course materials	−.48	.20	−.09
III	Compare authors read in and out of the course	−.61*	−.71*	.21
IV	Express in writings critical judgments on problems relative to those in the course	−.42	−.71*	−.03

* Significant at the 5 per cent level of confidence.

Part I, which requires only the recognition of correct factual alternatives, is the only one of the four which correlates positively with anxiety-readiness level. Part III, calling for independent preparation of materials prior to organizing and interpreting them on the examination, correlates negatively and significantly with the anxiety ratings. The coefficients for Parts II and IV indicate that when called upon to interpret the general views of a particular writer, or to treat narrow topics in a defined manner, those students whose anxiety is easily elicited perform less ably. The obtained results thus bear out predictions based upon an a priori classification of examination materials.

In summary, individual susceptibility to anxiety varies inversely with the ability to reorganize or to make fresh attacks on familiar topics. It would appear that high anxiety has deleterious effects upon the ability to *improvise or change* set.

D. THE CONSCIOUS THOUGHTS

It was previously hypothesized that a relationship may exist between the amount of thinking the individual does about himself and his susceptibility to anxiety. A rank-order correlation of .65 (significant at the 5 per cent level) was obtained between the Source of Thought classification *Self* and anxiety threshold. The frequency of those thoughts grouped under *Concern about the Self* correlated .47 with anxiety. These coefficients suggest that the anxious individual is troubled with feelings of personal inadequacy. Consequently, we should predict that anxiety will vary directly with negatively directed thoughts and self-confidence. To test this, the classification *Concern about the Self* (See Appendix A) was divided into "positive" content units (e.g., personal adequacy, feelings of intellectual superiority) and "negative" content units (e.g., those related to inferiority feelings).

The obtained correlations, —.33 with feelings of adequacy and .60 with feelings of inadequacy, support our initial hypothesis concerning the direction of thoughts about the self, and appear to indicate a withdrawal into the self.[8]

These results suggest that other kinds of withdrawal from the threatening class situation may also occur. Thus we might expect that the higher the anxiety readiness, the more the individual will attend to objects in his environment, rather than to the subject matter itself. The obtained rank correlation between anxiety level and the frequency of attention to irrelevant concrete objects in the classroom was .55, while the correlation with thoughts about the subject matter was only .10. Since the objects in the class are quite structured, definite, and unchallenging, it is possible that they become anchorage points lending themselves to remove the individual from the class material.

It is assumed that the individual confronted with a threatening situation fears his inability to cope with that situation, as well as future occurrences of similar threats. Consequently, it is plausible that the individual will attempt to forearm himself for a class discussion by looking through notes, reading the text, or computing answers in anticipation of being called on to recite. This formulation is supported by the positive correlation of .32 obtained between anxiety-readiness level and these kinds of class activities.

Since the ability to deal with new materials and organizational problems is somewhat suppressed by high anxiety, little attempt to relate old class materials to new subject matter is expected. The relevant correlation of $-.70$ was significant at beyond the 5 per cent level.

---

[8] The first class session recorded was oriented towards individual performance on an assigned essay. Students with varying levels of anxiety readiness were concerned with their grades and performance. The following excerpts are selected from the interviews with students having high, average, and low anxiety readiness:

High Anxiety Readiness

Ernest: "I was concerned about my essay. I had quite a terrible grade on it. I wondered whether or not I should speak to him [the instructor] after class, or call him at home this evening. I was also interested in the others' grades."

Average Anxiety Readiness

Robert: "My primary concern was my own paper. I looked at it; I was apprehensive as to what I should do. I was really very worried as whether or not I had gotten my idea in writing. This was intensified by his example, about how this girl did put her thoughts down. I was confused as to his grading system. I didn't know how well I did even after I got my paper back."

Jack: "I thought of my essay. It seemed inadequate to him, therefore, later to me. I was concerned with evaluation, and therefore, I listened to him. . . . I wondered who had written this essay he had read. I wondered if the other essay I wrote in a personal tone wouldn't have been better than this impersonal one that I handed in. . . . I wondered what Frances' grade was."

Low Anxiety Readiness

Frances: "I was thinking of my paper. I was concerned about it. I had seen it and wasn't enthused about it. It was rather short."

It will be remembered that the problem-solving category developed had been divided on an a priori basis into two categories: (*a*) the ability to give factual answers; and (*b*) the ability to relate and synthesize materials. Anxiety might be expected to correlate positively with the former kind of ability, since previous preparation would enable the student to make correct responses. It should be noted, however, that we were not concerned with the actual correctness of the reported answers, but were interested only in the efforts made in that direction. However, the obtained correlations, .44 with the whole category of problem solving, $-.47$ with application, and .24 with solutions, though suggestive, are not statistically significant.

In summary, the data obtained from the reported conscious thoughts indicate that if anxiety is aroused beyond the individual's threshold, responses which obtain tend to be stereotyped and readily available. When new approaches to familiar material or new materials are called for, response is made via familiar and habitual approaches, the anxious individual being less capable of improvising in the situation. In addition, withdrawal from the situation may take place in the form of concern with concrete objects in the class situation, or in terms of self-engrossment in the direction of concern with self-inadequacies.

E. Discomfort-Relief Quotient

The discussion thus far has been concerned with the correlates of anxiety level and the anticipatory nature of anxiety. It may be assumed that it is the poorer student who is most affected by anticipatory anxiety, and who may, consequently, tend to have a higher anxiety-readiness level. However, the scholastically superior student may be equally uncomfortable in the classroom situation, though for a different reason. He may be bored and impatient to proceed to a more challenging consideration of the subject matter. Whereas the discomfort of the poor student may result from anxiety due to feelings of inadequacy, the discomfort of the good student may result from lack of stimulation or boredom.

To test this hypothesis, the 11 subjects were divided into three groups on the basis of the final grades obtained in Social Science I: the *A-B* group, the *C* group, and the *D-F* group. As shown in Figure 1, the mean DRQ's for both the high and low groups are not significantly different. As anticipated, the good and the poor students are the ones who experience the greatest amount of class dissatisfaction.

When the mean DRQ's were computed for groups divided on the basis of performance on entrance examinations, the results obtained for both the Reading test and the ACE examination followed the pattern found with course grades (see Fig. 2). For the Writing test, however, DRQ varied inversely and in a linear, rather than the U-shaped fashion, with test performance. The average student for the group, while he may experience discomfort, is the one whose understanding is met somewhat by the class discussion. In other words, it is possible that the obtained relationships indicate that the class situation is geared to the middle student at the expense of losing the good and poor student by boredom for the one and difficulty for the other.

Finally, the DRQ correlated −.64 with ratings of anxiety-readiness level. This coefficient, significant at beyond the 5 per cent level, suggests that the individual may develop various mechanisms, such as withdrawal or repression, to prevent his anxiety from reaching consciousness.

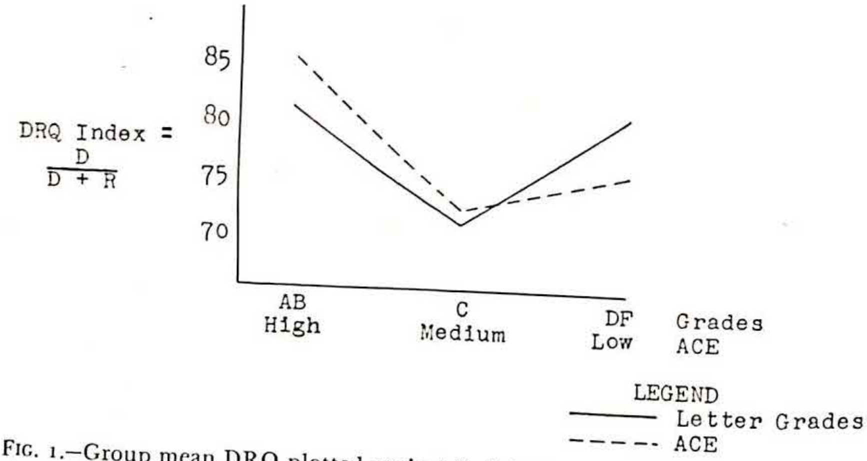


FIG. 1.—Group mean DRQ plotted against Social Science I grades and ACE score; groups High, Medium, and Low.

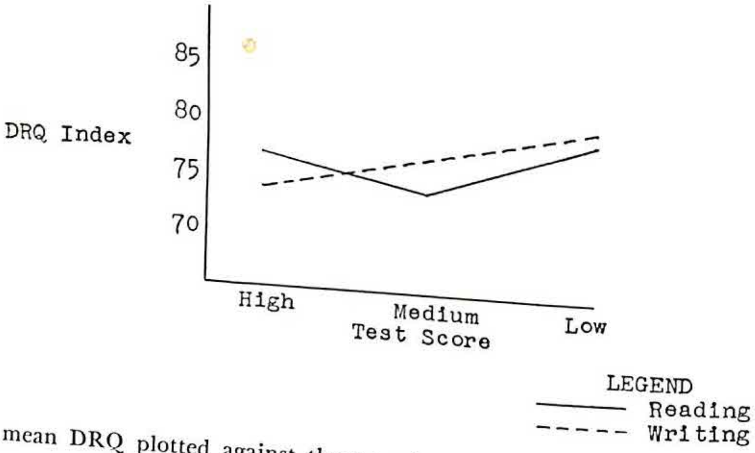


FIG. 2.—Group mean DRQ plotted against three performance groups in the Reading Achievement test and the Writing Mechanics test.

V. RIGIDITY IN THE LEARNING SITUATION

A. THE CONCEPT OF RIGIDITY

Common to the approaches to rigidity (8, 9, 15, 16, 17) is the idea that when the values of the individual are threatened or are especially vulnerable to attack, often because of environmental contradictions, he is relatively less able to adapt to a new situation. Intense emotional involvement and concreteness of behavior are observable correlates and possible preconditions of the occurrence of rigid behavior. Freud conceived rigidity as a defense mechanism aimed at preventing the rise to consciousness of certain repressed impulses which are unacceptable to the person's superego values.

Unexpected disturbances, if sufficiently severe, cause "catastrophic reactions" (9) which are repetitive and stereotyped by virtue of their dependence upon the individual's long-standing security patterns. Accordingly, the individual with a low "catastrophic threshold" tends to avoid those situations requiring spontaneity and variability of response. He seeks out and functions best in a highly organized, unchanging environment. In the light of this, we set up our three hypotheses for investigation here: Hypothesis I: *The greater the predisposition towards rigidity, the more likely will the individual fixate on a response, and not alter it, especially in the direction of the class material.* Hypothesis II: *The greater the predisposition to rigidity, the less able is the individual to organize material in a stress situation. He can, however, cope with structured and available concrete materials.* Hypothesis III: *In the general situation, the greater the individual's predisposition to rigidity, the more will he relate situational changes to himself.*

B. A MEASURE OF RIGIDITY PREDISPOSITION

On the Rorschach record, it has been noted (11, 12) that rigid subjects tend to minimize movement, color, shading, and perspective responses. Their responses are impersonalized, constricted, tend to be few in number, and are usually concentrated on the last few cards.

Responses showing good form ($F+$), but without shading, indicate poor ability to relate oneself satisfactorily to people and social institutions. Klopfer and Kelley (11) assert that the control of any subject of above-average intelligence whose record contains more than 50 per cent F, can be called constrictive. It may be assumed that the more closely the percentage of $F+$ responses approaches one hundred, the higher is the degree of rigidity.

The percentage of animal responses (A) appears to follow the general pattern of inhibition versus impulsiveness as revealed in the rest of the Rorschach record. In general, a high $A\%$ tends to indicate inhibition, rigidity, or a narrow range of interests.

In the present study, two Rorschach experts were asked to rank independently the 11 protocols along a continuum of predisposition towards rigidity. The obtained rank-order correlation of .75, computed as a measure of rater reliability, was significant at beyond the 5 per cent level. After the pooled ratings obtained in this manner had been assigned, the subjects were divided into high, average, and low groups on the basis of their apparent predisposition to rigidity.

C. THE TEST SITUATION

1. *The Aptitude Tests.* It has been postulated that the emotional reactions elicited in a test situation may tend to constrict the individual and interfere with his cognitive performance (15, 17). From our second hypothesis, we would predict a positive relationship between rigidity and the ability to deal with concrete materials. The low positive correlations which were obtained between rigidity on the one hand, and the ACE scores and the Reading and Writing tests on the other are presented in Table 1. Since each of these tests depends at least partially on recall without organization, the results are in line with our expectations. With only one exception, all of the coefficients obtained are low but positive.

2. *The Comprehensive Examination.* The parts of the comprehensive examination in Social Science I offer another area in which to test our second hypothesis. The obtained correlations between rigidity and performance on each of the four areas of the Comprehensive are presented in Table 2.

The index for Part I suggests that as

rigidity increases, so does the ability to deal with structured, familiar material in a stress situation. The value of rho (.73) is significant beyond the 5 per cent level. However, this increase may not continue indefinitely; for rigidity becomes maladaptive when extreme. Within limits, however, the above trend is quite probably effective. It should be noted that Part I calls merely for recognition of previously learned materials with little else required.

Part II deals with previously encountered materials, but in a new fashion. It may be assumed that some acquaintance with the manner of subject-matter treatment exists, and that the student has some ability to perform the necessary operations when he enters the test situation. The correlation here is not statistically significant.

The highly significant negative correlations between rigidity predisposition and the *organizational* parts of the examination indicate that rigid individuals lack the ability to use fresh attacks in dealing with such problems. It is possible that these subjects resort to methods proven successful in the past (i.e., "old learning"), when they find themselves unable to deal with the materials in the fashion demanded.

D. *Evidence of Rigidity in the Conscious Thoughts*

Experiences of perseveration of words and phrases in the classroom correlated with rigidity predisposition to an extent represented by a rho of .62, a value significant at the 5 per cent level. These students appear to be still concentrating on material the rest of the class has left, a phenomenon suggesting an inability to tolerate overlapping situations or to respond simultaneously to more than one set of stimuli. It may be that this "need for structure" carries over into the realms of problem solving and evaluative behavior.

Problem-oriented behavior was negatively though not significantly associated with rigidity predisposition (rho = −.14), and correlated positively (rho = .31) with orientation to the ideas under class consideration. It is possible that the more rigid individual tends to be receptive only to those ideas which are clearly defined and structured. Support for this hypothesis obtains from the correlation of .43 obtained between measured rigidity and the frequency with which objects within sight and hearing were perceived. External objects and concrete ideas share in common the attribute of tangibility; they call for little interpretation or projected meaning on the part of the individual. Though its verification must be left to future research, it is reasonable to suppose that the ideational responses of the "stimulus bound" individual may be as rigid and stereotyped as his behavioral responses—that rigid activity implies rigid thought.

Freud has postulated that in order to maintain his repressions, the rigid individual must set up an inflexible system of defenses. Any weakening of the defense mechanisms involves the danger that repressed, unacceptable tendencies may break through into consciousness. The necessity for continuously maintaining his defense system intact may promote in the individual an increasing rigidity and an increasing concern with himself (cf. Hypothesis III).

The interview data indicate that most of this self-directed thinking is concerned with the individuals' own inadequacies (see Table 3). It has been previously noted that the rigid person tends to take refuge in some kind of conformity, e.g., academic conformity. Thus the discrep-

TABLE 3

RANK CORRELATIONS BETWEEN RIGIDITY PREDISPOSITION AND THOUGHTS ABOUT THE SELF AND PEOPLE

Variable	Rho
Total Self	.60*
Self: Negative	.52
Self: Positive	.20
Teacher	−.55
Other students	−.20

* Significant at the 5 per cent level of confidence.

ancy between his social conformity drives and the realities of his own personality characteristics may cause to develop within the individual a negative reaction to himself, a state of affairs designated by Horney (10) as a "sacrificed subjective environment."

The data of the present study indicate that negative thoughts about the self are associated with increasing rigidity predisposition. It is possible that in turning his experiences towards himself, the rigid individual is able to turn away from those areas where his defenses are jeopardized, and thus remain within the narrowly circumscribed areas in which he feels secure.

E. Discomfort-Relief Quotient

By becoming more concerned with himself, the individual may psychologically withdraw from the objective situation, thus reducing threat to a minimum. The above results support the hypothesis that as the individual comes into a threatening situation, he will attempt either to withdraw or to introduce stability by clinging to concrete details and familiar objects. These affective elements are so fused into his perception that he becomes part of the external world which changes as these elements shift in *their dynamic relation to him.* If such mechanisms are effective, little discomfort should be experienced in the class situation, for the withdrawal and concrete perception will serve to protect the individual. As anticipated, none of the correlations obtained between rigidity predisposition and parts of the DRQ was statistically significant (Satisfaction = −.05, Dissatisfaction = −.43, Ambivalence = .49, Neutral = .18, and Total DRQ = .21).

If we can assume that rigid individuals tend to internalize threatening environmental stimuli, the negative correlation with dissatisfaction can be partially explained. It must be noted, however, that the DRQ represents a proportion between the amount of discomfort experienced and the sum of both idea units of satisfaction and dissatisfaction. Thus it may be that although the individual is physically removed from the situation, he is not entirely withdrawn from it emotionally, and accordingly continues to experience some discomfort. In line with the above, it may be noted that the correlation between rigidity and ambivalence is the highest of the values obtained for parts of the DRQ classification.

VI. MANIFESTATIONS OF NEGATIVISM IN THE LEARNING SITUATION

A. A Measure of Negativism

It is generally conceded that negativism is a learned response that may be aroused in the individual by stimuli perceived as threatening providing that such reactions are not contraindicated by the presence of contrary needs or drives. Rorschach (18) suggested that responses to white spaces (*S*) on the Rorschach cards are indicative of oppositional tendencies. In *S* responses, the relationship between figure and ground is reversed, the ink blot being seen as ground, and the white part of the card as figure. The 11 records obtained in the present study were rated for negativism by two Rorschach experts. The obtained reliability coefficient of .70, significant at the 5 per cent level, was deemed sufficiently high to justify assigning to each subject the combined rating of the two judges.

B. The Test Situation

1. *Aptitude Tests and Comprehensive Examination.* Tables 1 and 2 present the correlations of the negativism ratings with the aptitude-test scores and comprehensive-examination scores, respectively. Since individual negativism is quite removed and unrelated to the objective criteria that the College sets up to measure the students' ability, whether one scores high or low on the test should make little difference as compared with the amount of individual negativism. As might be expected, all of the obtained coefficients may be accounted for on the basis of chance. It may be hypothesized that the high motivation to succeed, characterizing a competitive situation, tended in this instance to prevent the actual expression of the latent negativism which may have existed. It is also possible that the College student accepts the comprehensive-examination program as part of the educational plan, and displaces his oppositional attitudes from the test situation to areas where grades and class status remain unthreatened.

C. Manifestations of Negativism in the Conscious Thoughts

1. *The Self as a Focal Point.* In the classroom situation, open rebellion against the procedure, the course, the instructor, or other class members is usually penalized in some manner. In such circumstances, the expression of resentment will usually remain covert, and may be directed toward the self. The obtained correlations between ratings of negativism and kinds of thinking about the self tend to bear out this hypothesis. Though not statistically significant, the coefficients suggest that thoughts about the self increase as the individual's oppositional tendencies increase (rho = .34), and that these thoughts tend to grow more self-critical as negativism grows more intense (rho = .52). Negativism and positively directed thoughts about the self are inversely related to a slight degree (rho = −.12). The thinking is mostly about the individual's personal and intellectual adequacy (or inadequacy) related to the other class members.

Moreover, positive thoughts about other people correlated −.45 with negativism, while negative thoughts about others correlated .67, a value significant at the 5 per cent level. In general, those individuals whose Rorschach records showed high negativism also displayed considerable concentration on the inadequacies of both the self and others in the class session, as the following interview excerpts indicate:

Ernest:

> "Frances . . . she's the smartest one. She is *so* smart. I don't feel prejudiced against them (the smart students) . . . not envious or anything like that. But I do wish that I would answer like that. I do respect them."
>
> "Miss X——— certainly does have her troubles, I thought. And that Max Factor makeup she uses. Heavens, she's practically a flaming siren. Then again, everyone looks at her, and she glances back so suspiciously."
>
> "I wasn't particularly interested in this problem. Instead, I thought of Mr. X——— [the instructor]. His teeth are so yellow; brown eyes with a beautiful parting in his hair. Golly, I wish he wouldn't wear those comical brown shoes . . . they look so cheap and have a rotten polish on them, probably liquid. Also, he stands with his back to the blackboard, so the back of his suit is full of chalk. The tie he wore wasn't half bad, either. Most amusing, I thought was

> the way he accentuates his problems . . . always underlining and putting circles around practically everything.
>
> "I felt good that Mark had more criticisms on his paper than I did. He didn't seem too bright to me . . . Stan, oh, he's so smart . . . it burns me up. But he has no moral values . . . runs around . . . for his capacity of IQ, he could rip the class wide open, which he doesn't do. I don't know why. He is so well educated, but he has such low moral values. He doesn't know how to live."

Mark:

> "His [Jack's] voice is irritating; the tone of his voice. Sounds like it was coming from deep down and could be very contrasting. When he talks in class, he keeps talking, and never says anything. I thought that Mr. X——— was going to get mad at him one of these days. He wasn't too bad yesterday, I thought. But a lot of times he starts to argue with Mr. X——— just for nothing, after the teacher has given him the right answer."

Jim:

> "[I was] thinking about where X——— gets his haircuts and his hats. Thought that they were the worst looking things I've ever seen in my life. He must get them at the Reynold's Club. Everyone who gets a haircut there . . . well, they're lousy."
>
> "I wondered why she didn't stay on the farm. I can't understand why these women don't stay at home and have babies. Thought that she typifies the typical U. of C. coed. Very intellectual and nothing else. I was thinking that this school had the reputation for very homely girls. See these girls walking around with these long, black stockings. Isn't that silly, Good Lord? I noted that Frances had a dress on yesterday. The first time since school began. I was thinking that we could always tell a U. of C. student when they came into our store."

2. *Negativism Directed Against the Subject Matter.* Direct argumentation and contradiction of subject-matter treatment in class may induce more anxiety than the amount of tension dissipated. Hence, we might expect that the student will show his resistance in class in covert ways which will not constitute too much threat for him. The obtained correlations between problem-oriented behavior and ratings of negativism are presented in Table 4. It is interesting to note that as disposition towards negativism increases, a decrease in active problem solving occurs. Moreover, negativism tends to correlate quite highly both with passivity (e.g., following along in the book, taking notes, waiting for new materials to be presented), and with evaluatory orientation (e.g., criticizing the relevancy and cogency of a point made, agreeing or disagreeing with the speaker). These data suggest that the negativism attendant upon poor problem-solving ability tends to be expressed either as dependence (cf. frustration → regression) or as irascibility (extrapunitiveness).

TABLE 4

RANK CORRELATIONS BETWEEN RATINGS ON NEGATIVISM AND PROBLEM-ORIENTED THINKING

Variable	Rho
Problem Solving	−.52
Active Participation (Evaluation)	.62*
Negative Evaluation	.34
Positive Evaluation	−.71*
Passive Participation	.54

* Significant at the 5 per cent level of confidence.

3. *Negativism Directed against Other Class Stimuli.* Cameron (2) contends that the individual may manifest negative attitudes by doing the opposite of what he believes demanded of him. Thus the individual who is negativistic toward the classroom situation may attend to irrelevant thoughts and stimuli rather than to the topics under discussion. Table 5 presents correlations between negativism and degree of irrelevant thinking and activity.

The correlation between oppositional tendencies and the performance of motor acts in class is significant at the 5 per cent level, suggesting that while *passive* in regard to the class material, the student can remain inattentive to it. The

TABLE 5

RANK CORRELATIONS BETWEEN NEGATIVISM AND THINKING AROUSED BY ENVIRONMENTAL STIMULI

Variable	Rho
Thoughts of concrete situations outside the class...............	.65*
Thoughts about physical activities..	.56
Thoughts about objects within sight.	.45

* Significant at the 5 per cent level of confidence.

highly significant correlation (rho = .80) between negativism and distractibility lends further support to this hypothesis. The index of distractibility used was the sum total of those ideas experienced during the class session which were *not* stimulated either by the topics actually discussed, by ideas directly related to those topics, or by the individuals' own feelings about them. These coefficients suggest that negativism which has been inhibited nevertheless tends to manifest itself both ideationally and behaviorally.

D. DISCOMFORT-RELIEF QUOTIENT

The initial class used in this investigation was recorded during the first week of attendance in the College. Materials and ideas at variance with the student's previous attitudes, values, and general orientation were presented. The following quotations, selected from the first recorded class, are presented to point up the climate being set up by the instructor:

> "Do I—do I pose the problem correctly? Now, I'm asking you, I'm asking your cooperation on this matter. I suppose a good number of you probably share the first part of the belief, and probably agree with the basis of government. . . . Is this argument a hard fact we've got to recognize as a consequence of our system? Or is there something wrong with the argument? I don't know. Did I make the argument clear? What's the answer?"

* * *

> "That makes a good deal of sense. . . . Any other questions on the questions? Well, Stan, you take it away now and tell me, let's talk this down. I'll try to be quiet; about—in twenty minutes—I'm going to come back in and see what we've learned and what we've got. Go ahead, Stan."

While these two excerpts appear to indicate a permissive and accepting classroom climate, the magnitude of the correlation between negativism and the DRQ for this class was quite high (rho = .57). This value was not even approximated for any of the other classes. But if this was the classroom climate, why was such a high coefficient obtained? It is possible that this result is attributable to the unfamiliar, "progressive," and perhaps threatening nature of the new material presented. As the course progressed, the students may have adapted to the "startling" character of the class work, or have been able to find some way of satisfactorily venting their negativistic feelings, thereby relieving their discomfort.

The trend of correlation values in time is not particularly consistent. Hence, it might be fruitful to inspect the individual records of the three students rated highest on negativism and attempt to seek out the dynamics of the qualitative shifts which took place during the period of investigation.

E. *Individual Patterns*

Table 6 contains indices for the three individuals rated highest on negativism as inferred from the Rorschach record. This table is of interest in pointing to the directions which individual negativism may take. Ernest ranked highest in the group of 11 subjects in orientation to people, the self, the instructor, and the class group, as well as next to the highest in thoughts about the class process. He received a *C* in the course, and his concern with class activity is low;

in fact, all of the values are almost equal. Mark, while concerned with activities distracting from the class topics, failed the course. Jim directed his opposition towards people and objects *removed* from the class situation, as well as to ideas related but not directly bearing on the immediate subject matter. In spite of this orientation, his course grade was a *C*. Thus, we can see three differing patterns, each unique for the individual, adding support to the trends of negativism discussed above.

TABLE 6

PERCENTAGE OF IDEA UNITS RANKED WITHIN EACH CATEGORY BY THE THREE MOST NEGATIVE SUBJECTS

Source of Thought	Ernest	Mark	Jim
Material under discussion	2.0	5.0	3.0
A related idea	2.5	6.5	11.0
Activity related to class (reading-writing)	2.5	11.0	2.5
Self	11.0	9.5	3.0
Instructor	11.0	4.0	1.5
Other students	11.0	3.5	8.0
Object within sight or hearing	7.0	11.0	2.5
Classroom situation or process	10.0	1.0	3.0
Someone not in the classroom	7.5	2.0	11.0
Object or situation outside the immediate environment	4.5	8.5	10.0

VII. SUMMARY AND CONCLUSIONS

This study has been concerned with investigating the relationships between individuals' recalled experiences in a series of class situations and certain relatively stabie personality characteristics as inferred from Rorschach protocols. More specifically, the following questions were investigated:

1. What is the relationship between selected personality characteristics (viz., anxiety, rigidity, negativism) and the kinds of thinking done by a group of college students in the learning situation?

2. What is the relationship between selected personality characteristics and products of learning, where such products can be inferred from test performance?

The majority of the data used in the present study were obtained by the *method of stimulated recall* recently introduced by Bloom (1). This involved recording 12 Social Science I class discussions in which the subject was a participant. Within a 48-hour period following the class meeting, each student was individually interviewed and the class recording was played back to him. During the interview, the recording was stopped at specific, preselected points (here called critical points) and the student was asked to recall the experiences he had during the actual situation. The ideas or thoughts reported during the series of weekly interviews were recorded verbatim and constituted the *conscious ideas* for this investigation. These data were collected through 64 interviews conducted over a four months' period with 11 students from one intact Social Science I class, registered for the first time in the College of the University of Chicago.

A. The unit of response classified was a statement, clause, or phrase denoting a complete idea; each such unit was termed an *idea unit*. The following system of classification for the reported conscious ideas was developed: (*a*) the stimulus or source of the experienced idea, called the *source of thought;* (*b*) the orientation or character of the idea unit, called the *orientation of thought;* and (*c*) the hedonic tone of the idea unit: this was a modification of the Discomfort-Relief Quotient (DRQ) developed by Dollard and Mowrer (5).

B. Performance on the Social Science I comprehensive examination was classified according to the type of required cognitive process called for on the examination. The Rorschach protocols were independently rated by two experts on anxiety readiness, rigidity predisposition, and negativism; in each case, a high score means *much* anxiety, rigidity, or negativism. The obtained reliability coefficients between judges' ratings based on the Rorschach were .80, .75, and .70, respectively.

A. Results

1. Anxiety readiness is negatively related to general level of performance on tests of scholastic aptitude. (Rho = −.37 with the ACE, −.66* with a reading test, and −.40 with a writing-mechanics test.)

2. When an analysis was made of the comprehensive examination, anxiety was found to relate positively only with the part calling for rote recall, rho = .34. Negative values were obtained with those parts calling, respectively, for analysis of familiar materials (rho = −.48), for comparisons of familiar and unfamiliar materials (rho = −.61*), and for critical evaluations of new materials (rho = −.42).

3. High anxiety readiness characterized those individuals who spent a great amount of time in class thinking about themselves in negative terms (rho = +.60*).

4. Rigidity predisposition, as inferred from the Rorschach, correlates positively with performance on materials calling for rote recall through identification (rho = .73*), but negatively with treatment of concepts or problems dissimilar to those previously studied (rho = −.71*), and those problems requiring new modes of attack (rho = −.71*).

5. In class rigidity predisposition is positively correlated with thinking about class stimuli most readily available, and with thinking about the self in negative terms (rho = .52).

6. Chance relationships were found between negativism and performance on different parts of the aptitude or achievement tests.

7. From the reported class experiences, evidence was found to indicate that negativism may be manifested in several ways. Values significant at the 5 per cent level were obtained between negativism and idea units concerned with negative and critical thoughts about other people (rho = .67*), and criticism of ideas expressed in class (rho = .62*). Negative thoughts about the self were highly, but not significantly, correlated with negativism (rho = .52).

B. Conclusions

The obtained results, while not definitive, strongly suggest the existence of relationships between certain personality, learning, and achievement variables. The following conclusions appear tentatively justified:

1. Anxiety and rigidity characterize those individuals less capable of improvising in a new problem situation. While problems calling for recall

---

* Significant at the 5 per cent level of confidence.

through recognition can be dealt with effectively, it is the organizational rather than the stereotyped behavior which is predominantly influenced by these personality variables.

2. Personality variables appear to be involved as influences in cognitive operations. These relationships, in the test situation, are demonstrable only when ability-test items are classified and analyzed *separately by type of required cognitive process*.

3. The obtained results carry implications for the measurement of personality through cognitive-ability tasks.

4. The conscious thoughts, as made available by the method of stimulated recall, open a vast new area for investigation. Research concerning the use of this technique is important in view of its potential application. Investigators in social psychology, group dynamics, and other fields will find the method a useful tool for research. The special merit of this methodology is that it permits the individual to participate freely in discussion or other social situations at one time, and to serve as an observer of his mental activity in the discussion at another time. Since the interviews are conducted after the completion of the social situation, they can have no effect on it. If future studies confirm our exploratory studies as to the accuracy and the detail with which the recall is stimulated by the recordings, this should prove a most useful method of inquiry in the social sciences. There has long been a need for a technique in the social sciences in which the act of investigation would have little or no effect on the situation being investigated. In general, the method of stimulated recall should be useful in many investigations of the relations between overt and observable behavior and the more subtle thought processes of which the individual is conscious.

The major result of the present investigation has been the demonstration that the conscious experiences of a group of students in a series of class sessions over a period of time, as well as performance on certain objective tests, are related to certain personality characteristics. The results are in accord with the view that no learning (with the possible exception of the conditioned-reflex type of learning) can be properly evaluated without regard for the personality of the learner. This is in sharp contrast to the early but persistent differentiation of behavior into cognition, conation, and affection, which are assumed to operate in virtual independence of each other.

REFERENCES

1. Bloom, B. S. The study of conscious thought processes by the method of stimulated recall. Publication pending.
2. Cameron, N. *A psychology of behavior disorders: A biosocial interpretation.* New York: Houghton Mifflin, 1947.
3. Claparède, E. La genese de l'hypothese. Étude experimentale. *Arch. Psychol.,* 1934, **24,** 1-155.
4. Cronbach, L. J. A validation design for qualitative studies of personality. *J. consult. Psychol.,* 1948, **12,** 365-374.
5. Dollard, J., & Mowrer, O. H. A measure of tension in written documents. *J. abnorm. soc. Psychol.,* 1947, **42,** 3-32.
6. Edwards, A. L. Statistical analysis for students in psychology and education. New York: Rinehart, 1946.
7. Freeman, G. L. *The energics of behavior.* Ithaca: Cornell University Press, 1948.
8. Frenkel-Brunswick, Else. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.,* 1949, **18,** 108-143.
9. Goldstein, K. *The organism: A holistic approach to biology.* New York: American Book Co., 1939.
10. Horney, Karen. *The neurotic personality of our time.* New York: Norton and Co., 1937.
11. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* New York: World Book, 1945.
12. Maslow, P. *Rorschach psychology.* New York: Brooklyn Coll. Press, 1945.
13. May, *R. The meaning of anxiety.* New York: Ronald, 1950.
14. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.,* 1939, **46,** 553-565.
15. Rokeach, M. Generalized rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.,* 1948, **43,** 259-278.
16. Rokeach, M. Prejudice and rigidity in children. *Amer. Psychologist,* 1948, **3,** 362. (Abstract)
17. Rokeach, M. The effect of preception time upon rigidity and concreteness of thinking. *J. exp. Psychol.,* 1950, **40,** 206-216.
18. Rorschach, H. *Psychodiagnostics.* (Trans. by B. P. Lemkau). Berne, Switzerland: Huber, 1942.
19. Titchener, E. B. *Lectures in the experimental psychology of the thought processes.* New York: Macmillan, 1909.

Accepted for publication, March 17, 1952

APPENDIX

CLASSIFICATION OF ORIENTATION OF THOUGHT[1]

I. PARTICIPATION

A. *Active Participation*[2]

Synthesis

9. Reconciliation of conflicting points
8. Relating discussion to general problem
7. Relating specific to some abstraction

Solution

6. Answer problems or questions raised

Application

5. Applies solution and/or idea—nonacademic
4. Applies solution and/or idea—other academic material
3. Applies solution and/or idea—in present course material
2. A related question or problem is raised
1. Searching for relevant point in the text or notes

B. *Evaluation*

7. Evaluation of aesthetic qualities
6. Evaluates the complexity and meaningfulness of the question or answer
5. Evaluation of the class process or procedure
4. Evaluated accuracy or authenticity of point or illustration
3. Evaluates relevance or cogence of point or illustration
2. Agrees with point or illustration
1. Disagrees with point or illustration
0. Surprised by point or illustration

C. *Comprehension*

7. Desires classification of a point
6. New insight or understanding: reorganization of his ideas on the subject
5. Recall of previously related experience or situations
4. Illustrates or relates it to another situation: nonacademic
3. Illustrates point or relates it to another situation: other academic material
2. Illustrates point or relates it to another situation: in course materials
1. Anticipates or wants a particular illustration given

D. *Passive Participation*

6. Curious as to what will be said
5. Simple understanding; following and listening with simple comprehension
4. Following and listening without comprehension
3. Following in book
2. Taking notes or following with book
1. Just following or listening; paying attention
0. Keeping ideas in mind, but also paying attention to other things

II. DIVERSIONS

A. *Individuals and Their Personal Characteristics*

8. Instructor's characteristics and mannerisms: approval, positive
7. Instructor's characteristics and mannerisms: disapproval, negative
6. Instructor's characteristics and mannerisms: nonevaluative and/or empathetic
5. Student speakers' characteristics and mannerisms: approval, positive
4. Student speakers' characteristics and mannerisms: disapproval, negative
3. Student speakers' characteristics and mannerisms: nonevaluative and/or empathetic
2. Other person's characteristics and mannerisms: approval, positive
1. Other person's characteristics and mannerisms: disapproval, negative
0. Other person's characteristics and mannerisms: nonevaluative and/or empathetic

B. *Ideas and Words*

3. Captivated or attracted by unusual word or phrase or idea or name
2. Amused by a remark
1. Annoyed or bothered by a remark
0. Tangential thoughts stimulated by word, phrase or idea

C. *Concern about the Self*

9. Desire of knowledge about self, e.g., grades
8. Concern about details of the self and activities; care of self
7. Feeling of personal adequacy: related to personal characteristics
6. Feeling of personal adequacy: superiority over someone in class, student or instructor, and related to discussion
5. Feeling of personal adequacy or positive attitude towards self: agreement with the speaker, and other people
4. Feeling of personal adequacy or positive at-

---

[1] The number placed before each category does not have any value except to facilitate ordering of the data.

[2] Involves the actual doing or effort, and not success of the action (such as giving the correct answer, accurate evaluation, etc.).

titude towards self: knowing the answer, comprehension or knowledge

3. Desires opportunity to demonstrate personal adequacy on a particular point by overt participation
2. Concern about possible demonstration of personal inadequacy: possibility of being called on to recite; wishes to avoid overt participation which will be judged
1. Concern about inadequacy of self: related to personal characteristics
0. Concern about inadequacy of self: related to discussion and subject matter presented

D. *Other Activities*

9. Tangential thought unrelated to the classroom stimuli
8. Not paying attention
7. Remembers the critical point, but cannot recall his thoughts of the time
6. Difficulty in hearing
5. Absent at this critical point
4. Looking at some object
3. Overt verbal participation in discussion
2. Speaking to someone in the class, not reciting
1. Occupied with other details: arranging books, clothing, getting organized
0. Reading

III. NO RECALL